Aalto University
School of Business

THE EFFECT OF MAINTENANCE CHARGES ON HOUSING PRICES IN FINLAND

Master's Thesis
Mikael Mäenpää & Niklas Mäenpää
Aalto University School of Business
Department of Finance
Spring 2021

Aalto University
School of Business

Aalto University, P.O. BOX 11000, 00076 AALTO
www.aalto.fi

Abstract of master's thesis

Author Mikael Mäenpää, Niklas Mäenpää	
Title of thesis The effect of maintenance charges on housing prices in Finland	
Degree Master of Science (Economics and Business Administration)	
Degree programme Master's Programme in Finance	
Thesis advisor(s) Elias Rantapuska	
Year of approval 2021	**Number of pages** 78 **Language** English

Abstract

A major part of the Finnish households' wealth lies on the housing market, and thus the valuation of dwellings concerns most Finnish people. There are an endless number of factors affecting the valuation of dwellings, of which the maintenance charge has been widely disregarded in the research. The relatively unique Finnish housing company system is the main cause of the ignoration. To our knowledge, we are the first to focus particularly on studying to what extent the maintenance charges are capitalized in the dwelling prices in our unique home market. This thesis has notable significance by improving the efficiency of the Finnish housing market. Efficient pricing of the dwellings reduces the spread between the bid and ask prices and eventually improves liquidity in the market.

We approach the research topic with the hedonic pricing method. Focus is set on Finnish dwelling transactions in apartment buildings and row houses between January 2000 and March 2021. The analysis is done using an extensive dataset provided by the Federation of Real Estate Agency.

Our main findings show that the maintenance charges are not capitalized to the dwelling prices in a significant extent as a one euro increase (decrease) in the monthly maintenance charge per square meter decreases (increases) the dwelling price per square meter by only 28 euros in apartment buildings and 22 euros in row houses on average. These numbers translate to extremely high implied discount rates of 43 % and 55 % when discounting to perpetuity, respectively. The analysis also shows that the high implied discount rates are mainly driven by the largest municipalities in which no capitalization of the maintenance charges on dwelling prices seem to appear. In practice, the results mean that the dwellings are overpriced with respect to the maintenance charges. Our results are probably driven by several behavioral biases, to which the housing market is particularly vulnerable to. One considering of buying a dwelling should carefully consider the level of the maintenance charges and focus on dwellings with low charges.

Keywords hedonic pricing, housing market, maintenance charge, housing valuation

Aalto-yliopisto
Kauppakorkeakoulu

Aalto-yliopisto, PL 11000, 00076 AALTO
www.aalto.fi

Maisterintutkinnon tutkielman tiivistelmä

Tekijä Mikael Mäenpää, Niklas Mäenpää

Työn nimi Hoitovastikkeiden vaikutus asuntojen hintoihin Suomessa

Tutkinto Kauppatieteiden maisteri

Koulutusohjelma Rahoitus

Työn ohjaaja(t) Elias Rantapuska

Hyväksymisvuosi 2021 **Sivumäärä** 78 **Kieli** englanti

Tiivistelmä

Merkittävä osa suomalaisten kotitalouksien varallisuudesta on keskittynyt asuntomarkkinalle, joten asuntojen hinnoittelu koskee useimpia suomalaisia. Asuntojen hintoihin vaikuttavia tekijöitä on lukematon määrä, joista erityisesti hoitovastike on jätetty kirjallisuudessa varsin huomiotta, pääasiassa siksi, että suomalainen vastikejärjestelmä on hyvin uniikki. Käsityksemme mukaan olemme ensimmäisiä, jotka tutkivat nimenomaisesti, miten hoitovastikkeet hinnoitellaan asuntojen hintoihin kotimarkkinallamme. Tällä opinnäytteellä on merkittävää arvoa, sillä se parantaa Suomen asuntomarkkinan tehokkuutta. Tehokas hinnoittelu vähentää kysynnän ja tarjonnan välistä kuilua ja parantaa näin markkinan likviditeettiä.

Käytämme hedonista hinnoittelumallia tutkimuksessamme. Käytössämme olevan datan tarjoaa Kiinteistönvälitysalan Keskusliitto. Analyysimme keskittyy suomalaisten kerros- ja rivitalo-osakehuoneistojen kauppoihin aikavälillä tammikuu 2000 – Maaliskuu 2021.

Päätuloksiemme mukaan hoitovastikkeita ei hinnoitella asuntojen hintoihin merkittävissä määrin. Yhden euron nousu (lasku) kuukausittaisessa hoitovastikkeessa per neliömetri laskee (nostaa) kerrostaloasunnon neliöhintaa vain 28:lla eurolla ja vastaavasti rivitaloasunnon neliöhintaa vain 22:lla eurolla keskimäärin koko Suomessa. Nämä luvut implikoivat noin 43 % ja 55 % diskonttokorkoja hoitovastikkeelle, kun diskonttoperiodi on ääretön. Huomaamme kuitenkin, että koko Suomen tuloksia selittävät vahvasti suurimmat suomalaiset kaupungit, joissa havaitsemme, että hoitovastikkeilla ei ole lainkaan vaikutusta asuntojen hintoihin. Toisin sanoen, voidaan sanoa, että asunnot ovat ylihinnoiteltuja suhteessa niiden hoitovastikkeisiin. Havaitsemiamme tuloksia selittävät todennäköisesti useat ihmisten käyttäytymiseen liittyvät epärationaaliset tekijät, joiden vaikutus korostuu erityisesti asuntomarkkinalla. Voimme todeta, että asunnon ostoa harkitsevan on syytä kiinnittää erityistä huomiota hoitovastikkeen suuruuteen ja mahdollisuuksien mukaan ostaa asunto pienellä vastikkeella puhtaasti taloudellisesta näkökulmasta tarkasteltuna.

Avainsanat hedoninen hinnoittelu, asuntomarkkina, hoitovastike, asuntojen hinnoittelu

Table of Contents

List of figures

List of tables

List of Appendices

1 Introduction

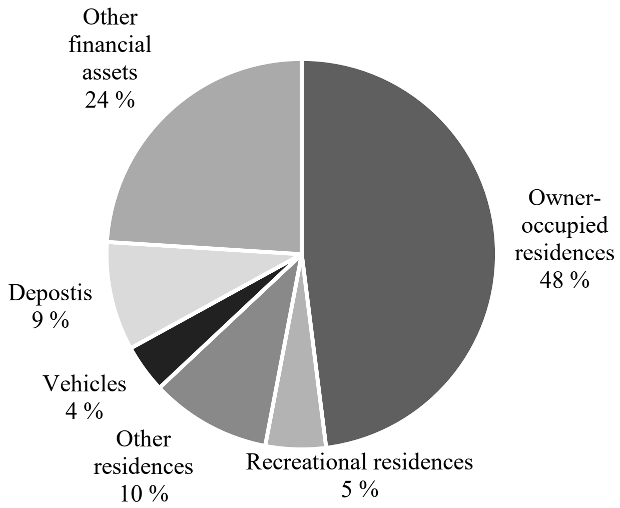
1.1 Background and motivation

The Finnish household wealth is very much clustered to real estate as can be seen from Figure 1. 63 % of Finnish household wealth was invested in real estate in 2019. The price development of real estate has been rather high in the past 20 years in the Helsinki metropolitan area (later referred as HMA). Other areas of Finland have seen much slower rise in real estate prices. Traditionally Finnish households have been rather cautious investors and at the same time owning your own house has been seen as a virtue in the Finnish society. Due to these trends, real estate is still where the Finnish wealth lies heavily, despite the recent increase in other forms of investing. These trends have resulted in a rather skewed distribution of wealth among Finnish households. 48 % of Finnish household wealth is invested in owner-occupied housing and another 15 % to other residences. 65 % of Finnish households own a dwelling that they use for own living purposes. 14 % of households own recreational residences and 16 % other residences. (Statistics Finland.)

There are also other interesting factors included in the housing market in Finland. First of all, there is a personal and often emotional element related to buying or selling a home. Many times, a home has sentimental value that is hard to justify rationally. Buying or selling one's home cannot be seen as a pure investment decision and the purchase price is affected by the seller's and buyer's emotions. What furthermore affects the pricing of real estate is that transactions are often made amongst non-professionals that do not have the necessary skills to value dwellings. Around 50 % of dwellings are held by private owners, instead of professional investors, for their own living purposes as can be seen from Figure 5. Many behavioral biases also affect the valuation of housing. Purchase decisions are bound to be irrational in the traditional asset valuation context. All these factors contribute to the fact that housing markets are not fully efficient, and mispricing occurs. When combining these factors to the fact that Finns have invested so heavily in real estate, we have an intriguing market setting for analysis. The importance of real estate prices on the Finnish household's wealth and consequently to the whole Finnish economy makes the valuation of real estate in Finland an interesting topic.

Figure 1 – Distribution of Finnish household wealth in 2019

Figure 1 shows the distribution of Finnish household gross wealth in 2019 divided in selected wealth groups. Real estate is separated in owner-occupied residences, recreational residences and other residences including investment real estate. Other financial assets include publicly listed and unlisted shares, retirement and other insurance investments, forest, land and other financial assets.



Source: Statistics Finland, household wealth 2019.

The Finnish housing market is divided in single-family houses and dwellings in apartment buildings and row houses. The scale of other forms of housing compared to the forementioned is relatively minor. The owner fully accounts for the costs related to maintaining or upgrading a single-family home. In apartment buildings and row houses the costs of maintenance and upgrades can be divided in two components: common costs for the whole building and costs related to a specific dwelling. The common costs are often substantially higher than costs needed to maintain a specific dwelling. In addition, the owner of the dwelling can decide not to maintain it, but the common areas and structural parts of the building need to be maintained to assure habitability of the dwellings. The common costs include maintenance of common areas and structures, heating costs, cleaning of the common areas and land rents to name a few. Limited liability housing company owns the apartment building or rowhouse and the residents only have share ownership of this company that entitles in the possession of a certain dwelling. The housing company charges the residents in order to cover its maintenance expenses. We call this charge the maintenance charge in this thesis. The monthly maintenance charges were on average 4.07 euros per square meter in apartment buildings and 3.14 euros per square meter in row houses in 2019 according to Statistics Finland. These figures turn into approximately 50 euros and 40 euros on yearly levels, respectively. Dwelling prices typically range between 1 000 – 10 000 euros per square meter in Finland with average prices for old dwellings being around 2 000 euros and for new dwellings around 4 000 euros per square meter in 2019 (Statistics Finland). When comparing the yearly maintenance charges to typical dwelling prices it is easy to see that these maintenance charges account for a large portion of the overall costs of buying and owning a dwelling in the long run.

A rational buyer makes an offer for a dwelling after considering and valuing all the characteristics of the dwelling. Also, all future costs and income related to the purchase should be taken in consideration and discounted. There is little knowledge on as how much the difference in maintenance charges really affects dwelling prices. Much of the differences originate from conditional differences between the buildings. Poor condition naturally increases future maintenance needs. Knight, Miceli & Sirmans (2000) find that buyers of real estate do not take in account the full effect of future maintenance costs. When the building or surrounding areas belonging to the property are in bad condition, it is hard to estimate the future maintenance needs and thus the level of maintenance costs. On average, the maintenance costs are quite stable over time, however. In addition to monetary costs, maintenance often comes with non-monetary cost e.g. discomfort from renovation. A rational buyer should also discount all future opportunity costs related to the purchase of a dwelling.

In the literature, personal discount rates are typically found to be rather high among ordinary people. Benzion, Rapoport & Yagil (1989) and Thaler (1981) studied personal discount rates with different time periods and sums of money. Benzion et al. (1989) found that with small sums and short periods, the yearly discount rates can be as high as 60 % (1 year and $40), but with larger sums and longer waiting periods around 10 - 15 % (4 years and $5 000). Thaler (1981) found even higher discount rates but used smaller sums and shorter periods in his study. There is evidence of strong current moment bias as personal discount rates are often observed to decline as the waiting period increases (Benzion, Rapoport & Yagil, 1989; Thaler, 1981). The same people that possess these high personal discount rates are the ones buying homes for themselves. The market is also very vulnerable to other behavioral biases as it is dominated by nonprofessionals. The assumption of incorrect pricing of maintenance charges in Finland is reasonable and worth studying in more detail. The aim of this thesis is to investigate how maintenance charges are taken in account when making decisions about dwelling transactions. This thesis is of great value to anyone planning on making a transaction of a dwelling in either personal or investment use.

1.2 Research questions and contribution

The concept of net present value (NPV) was first formalized by an economist Irvin Fisher in 1907 in his pioneering paper *The rate of interest* (Fisher, 1907). To put it short, NPV tells the value of future cash flows as of today, taking into account the size, composition, probability and distribution of cash flows in time. In this thesis we often refer to this theory of NPV as capitalization theory, even if they could be argued not being the same thing precisely. The theory of NPV is extremely widely used in the field of finance and is applicable in our study as well. According to the theory, an increase (decrease) in the future maintenance charge should be reflected to the dwelling price negatively (positively) today. In other words, dwellings having higher maintenance charges should be valued lower today and vice versa, compared to otherwise similar dwellings. There is risk involved in the level of maintenance charges and secondly, money has time value due to the opportunity costs making future maintenance charges less valuable than today's charges. Our thesis constructs over this idea. We will not try to develop new theory in the pricing of maintenance charges; we only use the capitalization theory to examine the phenomenon.

Different housing characteristics affecting the price of a dwelling have been studied both internationally and locally very extensively. However, there is only a limited number of research papers examining housing maintenance costs internationally and we find no research papers paying particular attention to maintenance charges in the Finnish housing market context. Also, international studies have examined how the components of maintenance charges (heating, taxes etc.) affect housing prices, but not how the whole maintenance charges themselves affect the prices.

Similar limited liability housing company models as we have here in Finland are only used, according to our knowledge, in Netherlands, Norway and Austria. We find no previous literature on this matter from these countries. Thus, our study contributes to the existing literature by providing new insight on the effect of the maintenance charges on housing prices, and thereby provides knowledge that benefits most Finnish people, but the results can also be used to assess the market in these other three countries. Our results should benefit the market by making the pricing of dwellings more accurate and thus resulting in an increase in the liquidity of the market. According to our results dwellings seem to be overpriced w.r.t maintenance charges. Furthermore, according to the results one should buy dwellings with small maintenance charges. The overpricing seems to be strongest in the largest cities and especially in the HMA. Apartment buildings suffer from stronger overpricing than row houses in general. We compare our results to the international papers studying the components of maintenance charges to validate our findings. We find similar results, but the overpricing we observe seems to be stronger. In short, the maintenance charges are reflected on dwelling prices on some level, but the price effect varies greatly across the country.

The research questions defined in a more concise manner are:

1. What is the effect of maintenance charges on dwelling prices in Finland?
2. How large is the implied discount rate of maintenance charges?

Other interesting questions we aim to answer and their rationales:

3. Does the price effect differ across municipalities?
   3.1. Does the market hotness explain the possible differences across municipalities?
   3.2. Does the dwellings' hotness explain the possible differences across municipalities?

   We suspect that when the housing market is hot and purchases need to be made fast, buyers pay less attention to the maintenance charges.

4. Is the effect of maintenance charges on dwelling prices linear w.r.t the level of the charge?

   If the maintenance charge per square meter is abnormally low or high, it may draw the buyers' attention resulting in more accurate pricing of the maintenance charge.

5. Is the effect stronger in large apartments with larger absolute charges?

   People tend to discount smaller sums with larger discount rates than larger ones. Thus, changes in small maintenance charges might not be presented in the value of the dwelling fully.

1.3 Scope of the study

The study is focused on dwellings in apartment buildings and row houses in the time period between January 2000 – March 2021. Single-family homes are left outside of the scope since they are usually self-operated rather than by a housing company, and thus, there is no actual maintenance charge. Subsidized housing is also left outside of the scope of this study as well as transactions between real estate investment companies. The geographical scope covers all the Finnish municipalities, which enables us to make assumptions and comparisons between different areas of Finland that have specific characteristics.

1.4 Structure of the thesis

The rest of this thesis is structured as follows: section *2 – Market setting* covers the different characteristics of the Finnish housing market and present the maintenance charges in a more thorough fashion. Section *3 – Theoretical background* comprises theoretical background related to housing price formulation. In section *4 – Data* we dive deep in the data and in section *5 - Methodology* we present the methodology used in our thesis. In section *6 – Results* we present and analyze the results and disclose our interpretations of the observations. In section *7 – Discussion* we discuss potential explanations for our results and then conclude the thesis with our main findings and further research suggestions in section *8 – Conclusions*.

2 Market setting

2.1 Finnish housing markets

To understand the price formulation in the Finnish housing markets and why maintenance charges may not be taken in account in full, we need to take a quick peek in the characteristics of the market. Above all, housing pricing mechanisms and cost structures related to housing vary significantly between countries and legislations. Finnish housing market is special in nature. We have quite a unique housing company system not found in many countries, legislation that affects housing pricing in some extent and a strong social security system with significant housing allowances that shift the housing demand and supply compared to countries with less regulation and societal support. First, we present some key characteristics of the Finnish housing market in section *2.1.1 – Characteristics of the Finnish housing market* and then explain the unusual Finnish housing company model and maintenance charges in detain in section *2.1.2 – Housing company charges*.

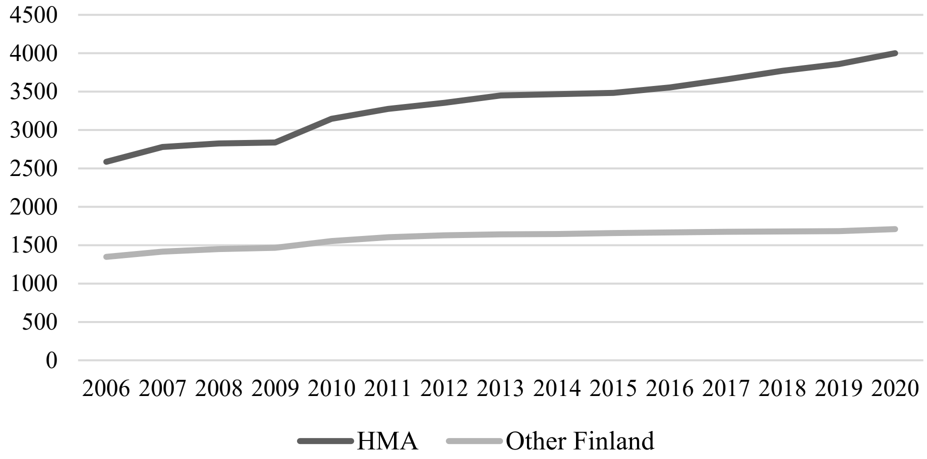
2.1.1 Characteristics of the Finnish housing market

According to Statistics Finland, in the end of year 2019 there were 3 076 000 dwellings in Finland in total. Of that number, 47 % were dwellings in apartment buildings and 14 % in row houses which both are subject to our study, while the rest 39 % were single-family homes. These are, however, nation-wide numbers, and the housing stock structure differs between the different areas of Finland. In larger municipalities where the population density is higher, apartment buildings are more common. For example, in Helsinki, apartment buildings accounted for 86 % of the housing stock alone in 2019. (Statistics Finland.)

Housing prices have fluctuated significantly over the years and in different areas of Finland as we can see in Figure 2. The price development has been particularly strong in the HMA, while the other parts of Finland lag far behind. Housing prices are formulated based on economic principles, and there are an endless number of factors affecting the supply and demand of the housing stock including people’s personal values, politics and legislation, the characteristics of the residences, interest rates and so on. Because there are numerous factors affecting the prices, they are difficult to forecast, and they may fluctuate unexpectedly.

Figure 2 – Average price per square meter of old dwellings in housing companies

Figure 2 represents the average dwelling prices per square meter in housing companies between the years 2006 and 2020 for old dwellings. Here the HMA consists of Helsinki, Espoo, Kauniainen and Vantaa. Municipality specific prices are weighted with the number of transactions in order to get the prices for the bundles.

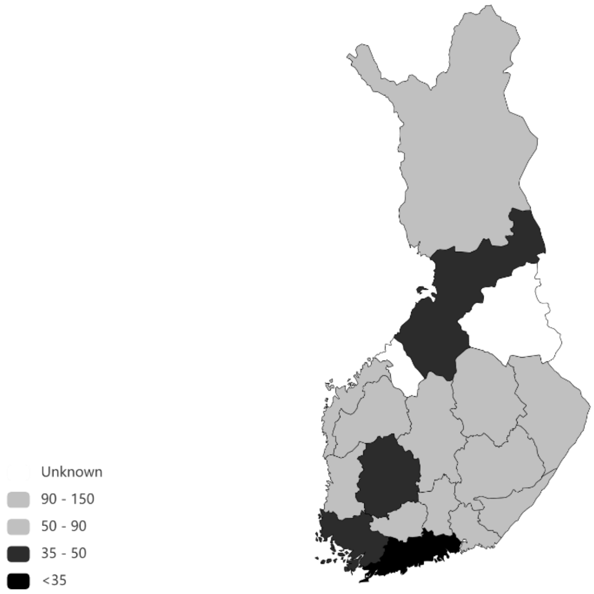


Source: Statistics Finland 2020, Average prices of old dwellings in housing companies and numbers of transactions by municipality, 2006-2020.

The price development is closely linked to the sales times. In hot market areas the housing prices are higher and the sales times shorter. In the Uusimaa region, the housing prices are the highest, and simultaneously, the sales times are the shortest. Figure 3 represents the sales times in more detail across the country.

Figure 3 – Sales times of old dwellings in housing companies by region

Figure 3 illustrates the sales times of old dwellings in housing companies by region between January 2019 and March 2021. Monthly sales times have been weighted by the number of dwellings on sale. Statistics Finland lacks data in two regions that are shown as white in the figure.

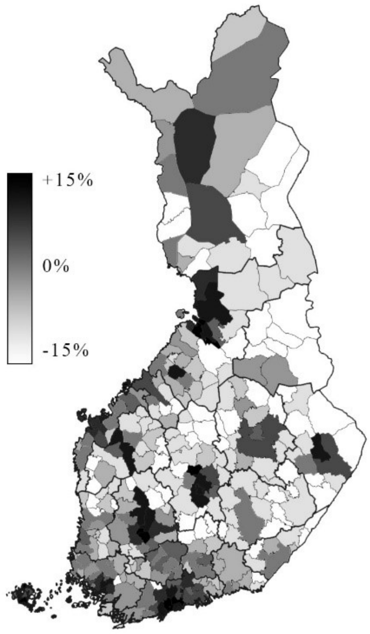


Source: Statistics Finland 2021, Number of dwelling advertisements and sales times by building type, region and month.

One notable factor affecting the housing prices is the development of population. The population grows especially in the largest municipalities and HMA. The population development is a sum of four components: net migration between the municipalities, immigration, as well as birth and mortality rates. To put it short, the faster the population growth is, the stronger the demand is with respect to the supply of housing. Figure 4 represents the development of population at municipality level.

Figure 4 – Development of population in relative terms at municipality level between 2007 and 2018

Figure 4 shows the development of total population in percentage terms at municipality level between 2007 and 2018. The population of black areas has grown over 15 % over the years whereas the white areas have had population growth of -15 % or lower.

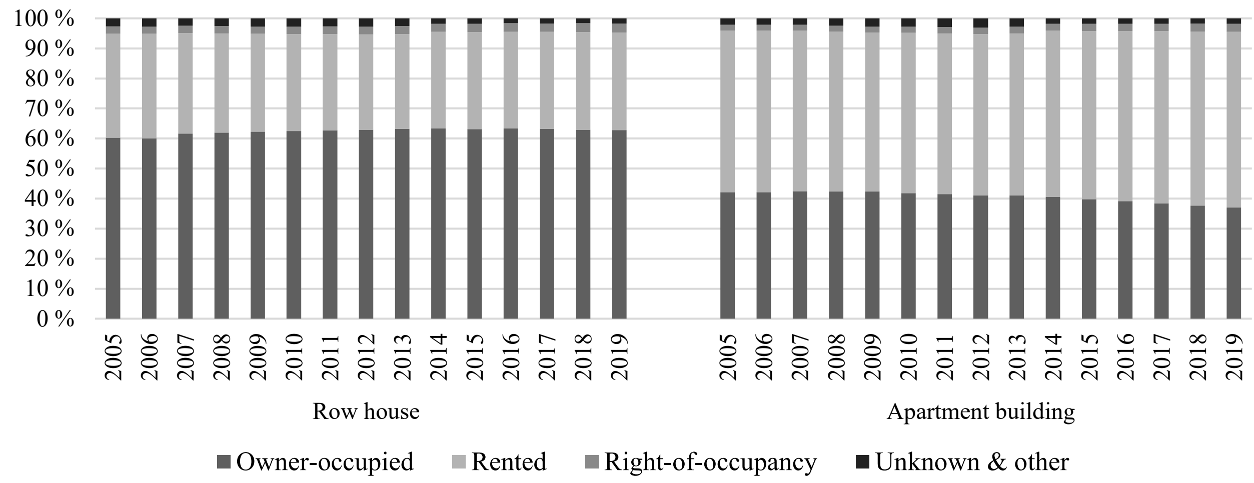


Source: Statistics Finland 2021, Vital statistics by month.

Dwellings can be separated into three main categories: owner-occupied, rented and right-of-occupancy dwellings. As the Figure 5 shows, the right-of-occupancy portion is particularly minor in both row houses and apartment buildings, while the distribution between owner-occupied and rented dwellings depends on the building type. There is not much fluctuation between the years. In larger municipalities such as Helsinki, the housing stock is more heavily skewed towards apartment buildings than on average in Finland.

Figure 5 – Dwelling units by tenure status and building type between 2005 and 2019

Figure 5 shows the distribution between tenure statuses by building type between the years 2005 and 2019 giving all households an equal weight regardless of the size of the household.



Source: Statistics Finland, 2020, Number of household-dwelling units and dwelling population by Year, Area, Type of building, Information and Tenure status.

2.1.2 Housing company charges

The Finnish ownership system of residential real estate is somewhat exceptional. The more usual system, condominium system, in which each dwelling is a property on its own, is used broadly in Europe and North and South America. The other, so called unitary system, which is used in Finland, is based on different principles. Here in Finland, a constructor sets up a limited liability housing company in order to start building a new apartment building or row house or essentially any dwelling to be sold later to residents and businesses. The ownership system is regulated by the Limited Liability Housing Company Act.

The constructor often finances the build with loans. The constructor starts selling the dwellings before the building is finished in hope that as soon as it finishes, all the apartments would be sold. The total debt-free price of a new-built dwelling includes the loan of the housing company that is allocated for the shares entitling to the possession of a certain dwelling. When the building is finished and all the shares sold to new dwelling owners, the constructor no more has right to the building. The housing company still has the loan that the constructor used to finance the build. When a dwelling is bought from the constructor, the buyer also becomes responsible of the housing company loan according to the portion of the shares of the housing company they own. The selling price of a dwelling is paid to the constructor, but the loan stays in the housing company.

The Limited Liability Housing Company Act determines the responsibilities of the housing company. A limited liability housing company operates in similar fashion as any other company. It has assets and liabilities as well as income and costs. The housing company collects payments called management charges (in Finnish *yhtiövastike*) from the residents in order to cover its liabilities and expenses. These charges are constructed of two different components: the payment of the housing company loan (capital charge, in Finnish *rahoitusvastike*) and the payment for the running costs of the housing company (maintenance charge, in Finnish *hoitovastike*). The capital charge is very similar to a mortgage in the resident's point of view and often these housing company loans are paid fully at the time of the purchase of the dwelling by the new resident with the mortgage they take. The Limited Liability Housing Company Act determines explicitly what expenses can be covered with a management charge; these expenses are accrued by:

(1) the acquisition and construction of the real estate;

(2) the use and maintenance of the real estate and buildings;

(3) the renovation and expansion of the real estate and buildings, and the acquisition of additional area (modernisation);

(4) the joint acquisition of a commodity related to the housing company's operations or the use of the real estate or building, and;

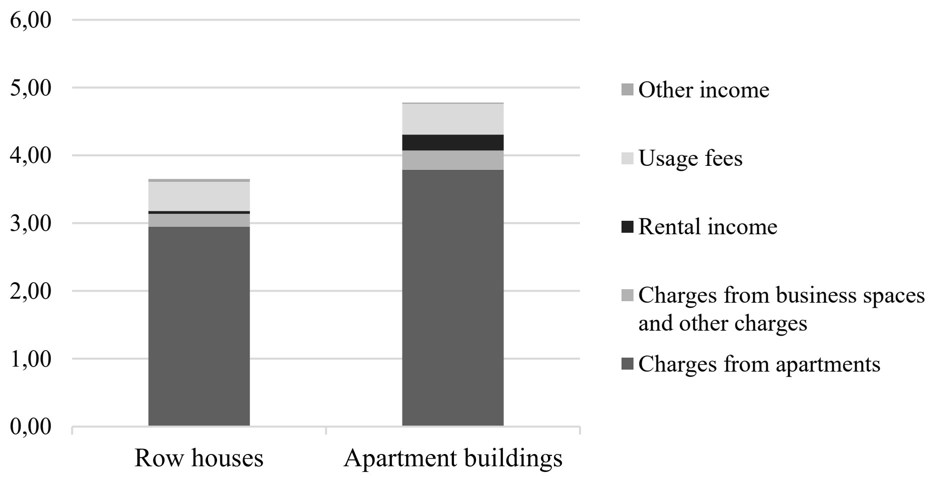
(5) the other responsibilities of the housing company.

(Limited Liability Housing Companies Act (1599/2009; amendments up to 547/2010 included; asunto-osakeyhtiölaki))

The maintenance charges usually do not cover the expenses of the housing company fully. The housing companies often also have other forms of income such as rents and other usage fees. It is quite common that the housing company owns the business spaces located in the apartment building and rents them. These spaces typically lie on the ground floor. In addition, housing companies may own individual apartments from the building and rent them. Usually, the housing companies also have parking spaces and common sauna departments that can be booked by the residents. The housing company charges these extra services by a separate fee that is not included in the maintenance charge. Figure 6 shows that in 2019 on average 85 % of the income of housing companies operating an apartment building came from maintenance charges. The corresponding figure for row houses was 86 %. The rest of the income came mainly from the usage fees and rental income.

Figure 6 – Housing company income by building type in Finland in 2019

Figure 6 shows the different categories of housing company monthly income per square meter in Finland in 2019. The data represents average income figures across whole Finland. Charges from business spaces and other spaces include one-time share-owner payments and repair and other special charges.

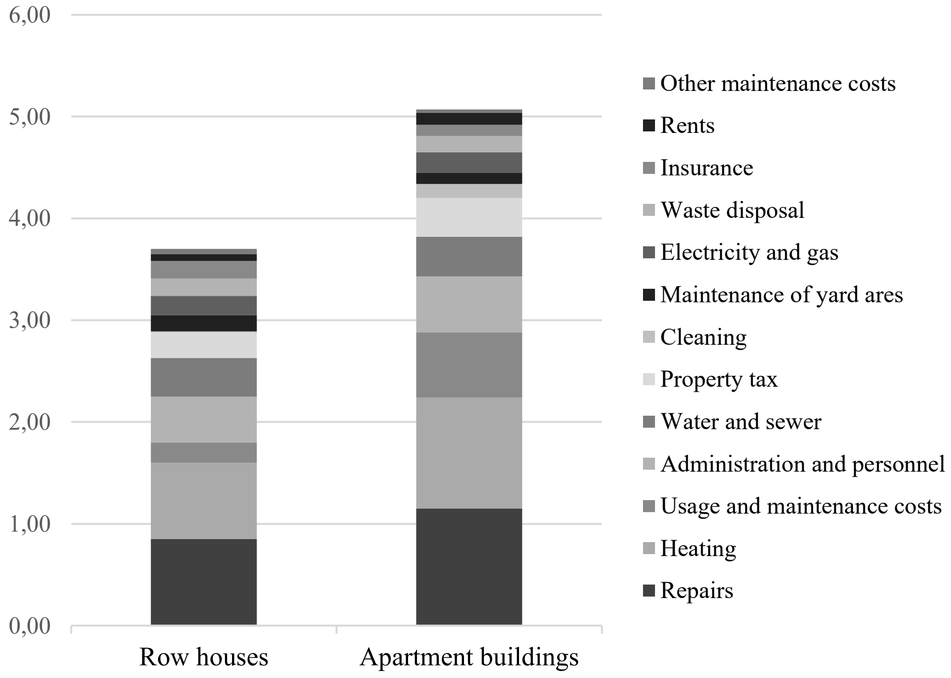


Source: Statistics Finland, financials of housing companies 2019.

The housing companies have the responsibility to maintain the building they operate. They have budget constraints and not everything can be always fixed or maintained properly. As can be seen from Figures 6 and 7, the total costs of housing companies were slightly higher than the total income in 2019 on average. The cost structure is quite extensive and often comprises of 10 or even more categories. However, only a few of these categories make up for most of the overall costs. According to Statistics Finland, in 2019 apartment buildings repair costs accounted for 23 % of the overall costs of housing companies, while heating accounted for 22 %, usage and maintenance for 13 % and administration for 11 %. The corresponding figures for row houses were repairs 23 %, heating 20 %, usage and maintenance 6 % and administration 12 %. Figure 7 also shows that other relevant cost were water and sewer, electricity and gas and property taxes.

Figure 7 – Housing companies' cost structure in Finland in 2019

Figure 7 presents the monthly cost structures of housing companies in Finland in 2019 for apartment buildings and row houses as euros per square meter. Total costs for row houses were 3.72 euros and for apartment buildings 5.07 euros per square meter.



Source: Statistics Finland, financials of housing companies 2019.

3 Theoretical background

3.1 How are housing prices formatted?

In the following chapters we dive in a bit deeper in the formation of housing prices in general. First in section *3.1 – How are housing prices formatted?* we discuss housing as a commodity, how the housing prices are formatted and often modelled, why submarkets inside the whole market matter and what characteristics specifically affect the housing prices. Then in section 3*.2 – Studies on housing prices and operating expenses* we look in previous literature on the matter.

3.1.1 Housing as a commodity

Housing stock is a private good as the owner alone can determine for the use of the property. However, the housing stock forms local neighborhoods that have also features of public goods. The properties of housing can be divided into two categories; the individual characteristics of a dwelling and the neighborhood attributes (Can 1992, 455). They both affect the value of the dwelling. O'Sullivan (1996) suggests that heterogeneity and the importance of the neighborhood distinguishes housing from majority of other goods. He also mentions immobility, durability, expensiveness, and high switching costs as similar kinds of distinguishing characteristics. As some of these characteristics play an important role in real estate pricing, we cover some of them in more detail.

Heterogeneity and immobility

The term real estate covers both, the actual physical building and the land the building sits on. The land and the buildings (including individual dwellings) can be sold together or separately. However, no area of land is identical to another area, and usually also buildings and dwellings are different from each other. When making a purchase decision, characteristics such as design, size, condition, layout of the dwellings and maintenance charges affect the attractiveness of the target. Buildings are inseparable from the land, and thus they are immobile. The land, in turn, is immobile by definition. The immobility plays a big role when choosing a dwelling, since the neighborhood characteristics cannot be later changed as easily as the condition or layout of the dwelling for example. Heterogeneity makes it extremely difficult to compare dwellings with each other for valuation purposes; there might be only a few comparable dwellings currently on sale or sold in the near past.

High switching/transaction costs

High transaction/switching costs in the housing market support the statement that the housing market is not efficient, and thus pricing errors might occur giving food for our research. Switching costs can be defined as one-time costs that the buyer faces when changing suppliers (Porter, 1979). Moving homes entails financial, mental and physical switching costs. Financial switching costs come in the forms of transfer tax (In Finland 2 % of the debt-free-price on dwellings in housing companies), capital gain tax (30 % or 34 % of the capital gain), and often also real estate broker's fee (often between 2 - 3 %). There are also other inconveniences in the moving process, such as the physical work required for moving. The old home may also have sentimental value as well. When moving to a new neighborhood or location the buyer also must adapt to the new neighborhood, local services and transportation and possibly is also separated from old friends and family. All these factors together make moving a rather expensive and cumbersome processes. Switching costs are a close relative to transaction costs that were initially introduced by Coase (1937). Transaction costs refer to the expenses that occur when making an economic transaction that are not accrued to any participant of the transaction. Gu and Hitt (2001) conclude that most research papers have found declined transaction costs to improve market efficiency.

Durability

Lifetime of a building depends highly on the level on maintenance performed over the years. Poorly maintained buildings and dwellings should trade cheaper than well maintained ones since they will have higher maintenance expenses in the future. High future maintenance expenses ultimately result in increases of the maintenance charges if they are not accounted way ahead. The stability of the maintenance charges over time affects the discount rate that should be used when valuing future maintenance charges. Later in our thesis we discuss on what would be a correct ballpark for the discount rate and compare our results to it. Both, the actual dwellings, and the common parts of the buildings need to be maintained. The Limited Liability Housing Company Act determines the responsibilities of the shareholder and the housing company.

According to the act, the owner of the dwelling accounts only for the maintenance of the interior. In a more precise manner, the Limited Liability Housing Company Act determines the responsibilities as follows:

> *"(1) The housing company shall be responsible for maintenance that is not the responsibility of the shareholder.*

(2) The housing company shall keep the building structures and insulating materials of owner apartments in good condition. Moreover, the housing company shall be responsible for the maintenance of systems for heating, electricity, data communications, gas, water, sewer and ventilation and other similar basic utility systems. However, the housing company shall not be responsible for sinks, tubs, bowls, basins or other such products located in the owner apartments. The housing company shall repair those indoor parts of the apartments that are damaged due to a failure in, or the repair of, the building structure or some other part of the building for whose maintenance the housing company is responsible.

(3) The responsibility referred to above, in subsection (2), applies to building structures, insulating material and basic utility systems that the housing company has installed or assumed responsibility for, and to the repair of the indoor parts of apartments to the current basic level within the housing company. The housing company shall also be responsible for any installations carried out or commissioned by a shareholder that are comparable to measures that the housing company has carried out or assumed responsibility for and the implementation of which the housing company has been able to monitor in accordance with this Act.

(4) The housing company shall also keep the building facade in good condition, including the part on balconies that are in the possession of shareholders, in accordance with chapter 1, section 3."

The breakdown of the responsibilities is particularly important in the context of this study since the running maintenance expenses of the building are included in the maintenance charge. The structure of the charges was covered in more detail in the section *2.1.2 – Housing company charges*.

3.1.2 Housing market dynamics

In order to understand the real estate valuation and the factors that determine dwelling prices, one must understand the general principles and dynamics of the housing market. One must also understand the cyclicality of the market, which makes the valuation of individual dwellings hard over time. The four-quadrant model that explains the dependencies between the real estate asset prices, rent levels, new construction and total stock area, was first introduced by DiPasquale & Wheaton (1992). The model separates the market into space and asset markets. According to the model, when for example new people move to the town and demand for space increases, the rent levels increase, which leads to increased property prices. Increased property prices make new construction projects more profitable and eventually, the total stock area meets the demand in the long-term equilibrium. The four-quadrant model has, however, some flaws. According to Colwell (2002), one of the main flaws is that the four-quadrant model does not show what happens in the protracted adjustment process. Because the housing stock adjusts particularly sluggishly to the new demand for space, the market often overshoots in terms of rent levels and property prices in short term before they meet the long-term equilibrium. The overreactions affect to the cyclicality of the market. From time to time, the real estate market might be over or undervalued compared to the long-term equilibrium. For example, Oikarinen (2007) argues that in Q2/2007 the housing prices were some 8 % over the long-term equilibrium in Helsinki. Inter alia for this reason, we control for the time fixed effects on postal code level in our regression models, as discussed in more detail in the section *5 - Methodology*.

3.1.3 Economic valuation approaches

There are plenty of economic valuation methods used in the housing market context. The choice of the method depends highly on the research question. In this study, our focus is on the prices of individual dwelling characteristics, especially on the maintenance charge. Since housing is a composite commodity, the are no markets and market prices for individual characteristics such as room number or the condition of the dwelling or the air quality. The literature generally distinguishes the methods that can be used to value individual characteristics of housing into two groups. They are the stated preferences-based and revealed preferences-based methods.

The first one, stated preferences-based methods, refers to methods that use information from hypothetical markets. They could be divided into subgroups, but generally speaking, these methods require survey respondents to either estimate how much they would be willing to pay for each characteristic, or alternatively, choose between different hypothetical scenarios (i.e. “would you prefer to have a sauna, a balcony or a more efficient heating solution”) (Carson & Hanemann 2005; Molin, 2011; McConnel & Walls 2005). However, these stated preferences methods have some flaws. Firstly, as McConnel et al. (2005) argue, people might have difficulties to value their preferences based on hypothetical markets. Secondly, gathering some comprehensive and high-quality survey data is particularly laborious.

Neither of these flaws are major concerns in the revealed preferences-based methods, that are used often in housing valuation studies, and consequently, also in this thesis. These methods utilize information of the actual, real choices made in the market. One of the most used revealed preferences-based methods in the housing market valuation is the hedonic model, which can be used to decompose the housing prices into individual factors.

3.1.4 Hedonic pricing method

Hedonic analysis is used to find prices for attributes within a certain product when we do not have precise knowledge of what attributes are bought but have clear understanding of the overall cost of the product. Hedonic analysis is done with a hedonic pricing model and is often used in situations where interest is on a heterogenous stock of goods. Lancaster (1966) was one of the first to model goods as a bundle of characteristics that form the actual heterogenous product. When markets are constructed of products that in fact are packages of different characteristics, such as the housing market explained in section *3.1.1 - Housing as a commodity*, observed market prices are comparable when the characteristic bundles are broken down (Rosen, 1974). Every dwelling is virtually different as they are built upon a different mixture of characteristics, that according to the theory, add/delete value of the dwelling. The price differences between dwellings only equalize the different characteristic bundles between the dwellings (Rosen, 1976). The hedonic model simply takes in account both, the individual physical characteristics of the dwelling and the neighborhood and locational attributes (Can 1992, 456). The idea of hedonic modelling is that the markets on which dwellings are sold reveal the shadow prices for the characteristics of the dwellings. The relationship between the characteristics and their effects on the pricing of the dwelling can be non-linear or non-monotonous, as we see later particularly with the age of the dwelling in section *5.2 – Determining the form of the variables*.

It is worth mentioning that the marginal/shadow prices of the characteristics of dwellings may not be found with the hedonic modelling. The supply of housing is fixed in the short term and people searching homes must do tradeoffs when choosing from the available dwellings at the time of the purchase. The available dwellings may have very different characteristics compared to the hopes/needs of the buyers if they had free choice over the characteristics and thus the prices of the characteristic observed merely tell about their relative value for the buyers. One reason for us using a large data across a longer period is that it helps revealing the effects better as the market has time to adjust, and thus could be assumed to being relatively close to the long-term equilibrium revealing the shadow prices.

The consumers operating in the housing market are also heterogenous and value different characteristics differently. Not every transaction in the housing market makes sense in an economic context, as there are many individual factors that affect the price formation. These individual and often very personal consumer specific factors are impossible to fit in the hedonic model and something is always omitted. There are of course also non-consumer buyers and sellers in the market, but for the most part the transactions are done between consumers. Furthermore, real estate investors and owner-occupiers may have very different preferences that will affect the prices that buyers are willing to pay for the characteristics of dwellings.

There is no general consensus about the form of the hedonic regression model among the researchers (see e.g. Halvorsen & Pollakowski 1981), and the explanatory variables can be categorized in multiple ways. Malpezzi (2003) admirably points out, that hedonic model specification is as much art as it is science. We use one common form of the model in which the explanatory variables are categorized into structural, neighborhood and locational characteristics.

Thus, the general form of our hedonic regression model is as follows:

$$P = f(S, N, L, T) \tag{1}$$

in which: P = debt-free price, S = structural characteristics, N= neighborhood characteristics, L = location, and T = time

3.1.5 Regional submarkets

It has been recognized already decades ago in the academics that the housing markets consist of multiple submarkets rather than being homogenous (see Grigsby 1963, Straszheim 1975). The idea of submarkets is that the dwellings within a certain submarket are close substitutes to each other. The supply of the housing characteristics as well as the value given to the characteristics may vary across the submarkets. For example, houses that share some specific characteristic can form a submarket even though they are physically located far away from each other. Some buyers for example, may have particularly strong preference to live close to water, no matter if it was in Helsinki or Tampere.

If there is non-explainable variation inside the market, it makes the regression analysis imprecise. Thus, the determination of the submarkets is crucial in order to find reliable results. However, it is challenging to find the factors that differentiate the individual submarkets. Parr (2007) suggests that the submarkets can be defined by four different ways; based on the extent of the built area, the area in which most of the consumption takes place, the area in which the people work and finally, the area from which the workforce is being drawn. These definitions are, however, difficult to implement, and thus, submarkets are often defined by means of district-specific indicators such as postal codes even though the district-specific definition is not always the optimal way to go; the postal codes are only artificial boundaries between the areas. Definition of the submarkets is crucial for our analysis and we use the typical postal code submarket definition in all of our regressions with the help of postal code dummies. There are also other forms of submarket distinctions we use, and one of them is the use of locational market "hotness" (dwelling price development in the area) as defining element to separate submarkets in one of our analyses.

3.1.6 Factors affecting housing prices

Before building a hedonic pricing model we should have a clear understanding of what variables should be used in our regressions. There are many problems related to obtaining the relevant variables that must be taken into account. The econometric conditions must be met, and on the other hand, the availability of high-quality data is often limited. We discuss these topics in more detail in sections *4 – Data* and *5 – Methodology* and focus here on the regression variables commonly used in the literature.

It is easy to get a sufficient understanding of the most commonly used variables from the extensive literature. The forementioned problems affect the choice and availability of the variables, however. As mentioned, there is no clear one solution fits all -type of model, but some variables tend to pop up in the literature more than others. The amount of variability in the number of variables is considerable. For example, Palmquist (1984) uses 30 variables and Lineman (1981) uses only three in their studies. Variables should not be added just by the sake of adding. Instead, we believe that every variable should have an intuitive explanation of how it would affect the price of a dwelling. Sirmans, Macpherson & Zietz (2005) summarize 125 previous studies on housing pricing and construct a table of the commonly used variables in the previous literature. This conclusive summary gives us the starting point in building our model. The study of Sirmans et al. (2005) focuses on international previous studies, but there is little research on the topic from our Finnish home market. The most applicable study on the Finnish market is Laakso (1997).

Table 1 – Common attributes found in housing pricing studies

Table 1 presents the most commonly used variables in hedonic housing pricing studies summarized by Sirmans et al. (2005) from 125 previous independent studies.

Variable	# Of appearances	Positive	Negative	Not signifficant
Lot size	52	87 %	0 %	13 %
Ln (lot size)	12	75 %	0 %	25 %
Area	69	90 %	6 %	4 %
Ln (Area)	12	100 %	0 %	0 %
Brick	13	69 %	0 %	31 %
Age	78	9 %	81 %	10 %
# of stories	13	31 %	54 %	15 %
# of bathrooms	40	85 %	3 %	13 %
# of rooms	14	71 %	7 %	21 %
# of bedrooms	40	53 %	23 %	25 %
Full baths	37	84 %	3 %	14 %
Fireplace	57	75 %	5 %	19 %
Air-conditioning	37	92 %	3 %	5 %
Basement	21	71 %	5 %	24 %
Garage spaces	61	79 %	0 %	21 %
Deck	12	83 %	0 %	17 %
Pool	31	87 %	0 %	13 %
Distance	15	33 %	33 %	33 %
Time on market	18	6 %	44 %	50 %
Time trend	13	15 %	23 %	62 %

Intuitively, it is easy to tell the most important characteristics when buying a dwelling; room-number and floor area are characteristics that separate sufficient dwellings for a specific buyer from insufficient dwellings. The age of the dwelling is also intuitively a very important measure of the fit of the dwelling for the buyer’s purposes. It is no surprise that Sirmans et al. (2005) find these three variables to be the most commonly used in the literature. It is worth mentioning that most of the existing literature studies single-family homes and thus variables such as garage, swimming pool and fireplace seem to be quite often used. In our context such characteristics are quite scarce in the data. Still, the summary by Sirmans et al. (2005) gives us a good starting point with characteristics area, age, size, number of rooms and distance (or location in our model). Clearly this list is not comprehensive for usage with apartment buildings, as for example a variable for elevator or floor number is not presented, when undoubtedly, they must influence the dwelling prices. Usually also socioeconomic and physical aspects of the neighborhood and accessibility to services are included in the models.

In Finland not many studies of housing characteristics are made. Laakso's (1997) study on housing characteristics is the most relevant. He finds that the relative total price of a dwelling increases almost linearly up until the size of 100 square meters and then slows down, but again accelerates after 140 square meters. Price per square meter decreases with size. (Laakso, 1997.) Housing prices decrease quite steadily the older the dwelling gets according to Laakso (1997). According to his dummy models, bigger drops in price are observed with dwellings constructed in 1980, 1970 and 1940 compared to the control group. The bottom price of dwellings in HMA is reached with dwellings built in 1930's and 1940's. From 1930's onwards the prices start to appreciate again, and prices of dwellings built in 1900-1909 are found being 8 – 13 % higher than dwellings built in 1930's and 1940's. He tests also other models and finds similar results. (Laakso, 1997.) It should be noted that Laakso's (1997) study was conducted over 20 years ago and it is hard to tell if the effect is a result from the poor quality of 1930's and 1940's buildings (or other typical characteristics of buildings built in this time) or just purely due to 60-year-old buildings being in their worst condition regardless of the construction time. Also pipe repairs are usually done when dwellings are around 30 – 60 years old, which increases the price of the dwelling significantly. Other structural characteristics that Laakso (1997) uses are lot efficiency and ownership of the lot. Lot efficiency increases price and leased lot decreases the price compared to otherwise similar dwellings (Laakso, 1997). Our data has no indicator of lot efficiency, but information of the ownership of lot is available and used in our model. Interestingly, Laakso (1997) has not included the maintenance charge as a control variable.

In addition to structural characteristics, Laakso (1997) studies locational and neighborhood characteristics extensively. Vicinity to coast has strong positive effect on dwelling prices; the price of a dwelling is around 25 - 50 % higher at the coast than around 1 kilometer away, depending on the model used. We also employ a variable indicating vicinity of a shore. Laakso (1997) also finds some other housing characteristics that affect the values of dwellings, but we are limited by the data and cannot use exactly the same characteristics.

The buyer of the dwelling is maximizing their utility when choosing the characteristics that comprise the dwelling. The buyer has a budget constraint that prevents them from acquiring all the characteristics they desire. The investment to a specific characteristic is balanced by the marginal utility and the cost of acquiring the additional characteristic. In this case, the bid function for characteristics is concave. For example, the marginal utility from expanding form a 20-square-meter dwelling to a one having 30 square meters is larger than the marginal utility from an equal increase of area in a 90-square-meter dwelling. Keeping this in mind we have to make adjustments in the model; the form of the variables *Area*, *Room number* and *Age* of the dwelling has to be tested to see non-linearities and non-monotonousness. We discuss these tests in detail in section *5.2. – Determining the form of the variables*.

3.2 Studies on housing prices and operating expenses

We can hardly find studies on maintenance charges. As stated, Finland is one of the few countries having a maintenance charge system. We are unable to find previous literature focusing particularly to the effects of maintenance charges on dwelling prices, but the effect that different operational cost attributes have on dwelling prices has been studied earlier. We use these studies as a reference to see what kind of results could be expected.

Dinan & Miranowski (1986) studied the effects of heating efficiency improvements to housing prices. They found that on average a $1 decrease in annual heating costs increased the value of a single-family house by $11.63 in Des Moines, Iowa, where the study was conducted. The implicit discount rate for future heating costs found in their study is around 10 % when discounting to perpetuity.

Harjunen & Liski (2014) study the effect of heating choice to single-family housing prices in their working paper. They find that on average heating technology choice has a 6 % impact on the value of the dwelling between electric heating and district heating. The estimated euro-effect of having district heating compared to electric heating is 21 080 euros on average. They estimate the actual expected annual cost savings from district heating for a period of 25 years and find that the cost savings are capitalized with 2 % discount rate to the selling price of the house. This capitalization rate is very low, but also the discount period is quite short. Harjunen & Liski (2014) clarify that the discount rate found differs from most studies where individuals' discount rates are estimated, as they tend to be significantly higher (see e.g. Thaler (1981). (Harjunen & Liski, 2014.)

Longstreth, Coveney and Bowers (1985) also study the effect of energy costs on dwelling prices. They find that dwellings using more gas are sold cheaper than dwellings using less gas. Their findings indicate that with 3 % real interest rate, the price difference they find implies a discount period of 12.5 years. Using 25 years as a discount period they get an implied discount rate of 8,5 % for the future energy costs. (Longstreth, Coveney and Bowers, 1985.) The implied discount rate found by Longstreth, Coveney and Bowers (1985) differs quite much from the rate calculated by Harjunen & Liski (2014). The variation of the findings in the previous literature makes our study also interesting as presumptions are hard to make.

In turn, Kahn & Kok (2014) study the effect of green labeling in Californian residential real estate between 2007 - 2012. They record that the average annual cost saving from green labelling compared to non-green labelling in residential single-family housing is around $720. The price premium paid on green labelled houses in their study is $14 800 compared to similar houses that are non-green labelled. This translates to a simple payback period of 21 years or a discount rate of 4.9 % when discounting to perpetuity. With a 25-year discount period (same as Harjunen & Liski (2014)) the implied discount rate is 1.6 %, which is again extremely low.

Tyvimaa, Gibler & Zahirovic-Herbert (2015) study the effect of ground leases on apartment prices in Helsinki between 2005 - 2012. In theory, houses built on leased lots should be priced lower than houses on free-hold lots (following also Laakso's (1997) findings) as the tenant is expected to pay yearly maintenance charges that include the lease payments. The findings indicate that on average the price premium of apartments on free-hold lots is around 5 % compared to apartments on leased lots at the time of the sale. This premium is much smaller than the one found by Janssen (2003) (10 - 14 %) but gives support to the fact that operational costs are reflected to the value of dwellings at some level.

Janssen (2003) studies the effect of lot leases to apartment building valuation in Stockholm city center between 1992 - 1994. He finds that the price difference of a lease-hold lot compared to a free-hold lot is between 10 - 14 % of the predicted price for average lease-hold property. He presents that a 6 % implied discount rate for the lease payments can be calculated when discounting to perpetuity (Janssen, 2003). The author concludes that when valuing apartment buildings, the market takes in account the effect of a ground lease contract. (Janssen, 2003.) However, these transactions were made between real estate professionals and might not be entirely comparable to our results.

Elinder & Persson (2017) is another very interesting study that finds inconsistent and intriguing results of the capitalization of property taxes. They study the effects a property tax reform during 2006 - 2008 in Sweden. The property taxes were lowered especially in the highest priced segment of the market.

According to the capitalization theory, the dwelling prices should increase at the time of the tax reductions to reflect the NPV of the cost savings. The results of Elinder & Persson (2017) indicate zero effect on prices from a substantial tax reduction for most of the dwellings in the treatment group. If the supply of housing was perfectly elastic, this result would be expected. However, the supply is fixed in the short term. Price reactions are observed only among the top 1 % most expensive dwellings. The effect observed in this segment is substantial and statistically significant, but still only half of the magnitude of the theoretical effect if calculated using a 50-year discount period. If 25 years is used as a discount period to calculate the theoretical expected effect, the observed effect corresponds well in the highest priced 1 % segment. Other price segments in the treatment group show zero effect compared to the control group. (Elinder & Persson 2017.)

The mechanisms behind the observed results remain vague. Elinder & Persson (2017) argue that the significant capitalization in the highest priced 1 % segment could be observed due to the following: these dwellings are situated typically on areas where free land is scarce and no new construction can be made, the reform is salient in this segment and the buyers of the expensive homes are more financially literate to calculate the NPV of the cost savings. (Elinder & Persson 2017.) We take ideas from Elinder & Persson (2017) and test if the dwelling price has an impact on how the maintenance charges are taken in account. We also test how packed municipalities such as Helsinki differ from other less crowded ones.

On the other hand, Oates (1969) finds that an increase in property taxes reduce the dwelling values around 2/3 of what would be expected with full capitalization of future taxes (calculated with 5 % discount rate and discount period of 40 years) in his New Jersey based study.

The previous literature suggests that operational costs are capitalized to dwelling prices to some extent, but the amount of capitalization varies significantly from study to study. Based on the literature we would expect to see quite strong capitalization of the maintenance charges as well. However, we find rather opposing results.

4 Data

4.1 Overview of data

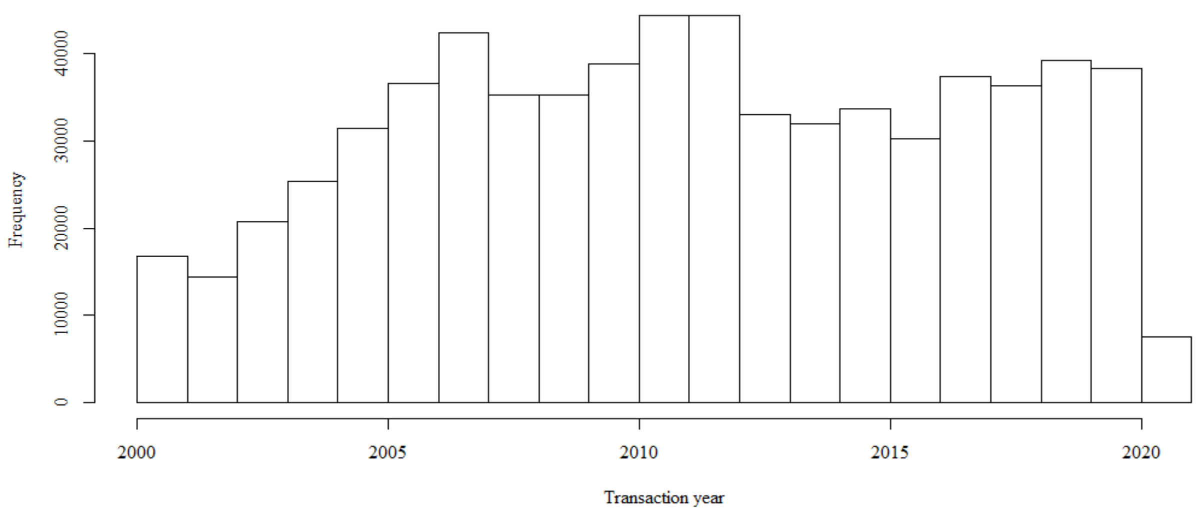
The main dataset used in our thesis is provided by Federation of Real Estate Agency (In Finnish *Kiinteistönvälitysalan Keskusliitto*, KVKL) and their database Hintaseurantapalvelu (later referred as HSP). HSP includes most major real estate agencies', such as Kiinteistömaailma, OP, Aktia, Remax, Realia Group and SP-koti's dwelling transactions totaling in 1.4 million transactions since year 1999. In addition, construction companies may purchase the service and add their transactions to the database.

The total amount of dwelling transactions in apartment buildings and row houses in our dataset withing the time period 01/2000 - 03/2021 is 1 035 774 observations. However, the data has gone through a comprehensive cleaning process after which we were left with 673 362 transactions to be used in our main regression model. The detailed data clean-up process is discussed in section *4.2 – Deriving the final dataset*. In addition to the transaction data, we draw some data from Statistics Finland's databases. Statistics Finland is a public authority producing statistics for open use. Of the 15 databases Statistics Finland offers free-of-charge, we use StatFin and Municipality key figures.

As Figure 8 shows, our dwelling transaction dataset is particularly skewed towards the later years, since over time new entities have joined KVKL and started sharing their transactions on HSP. For instance, one of the major agencies, Realia Group joined KVKL in 2005, and after that the number of annual observations has been rather stable.

Figure 8 – Histogram of the transaction year

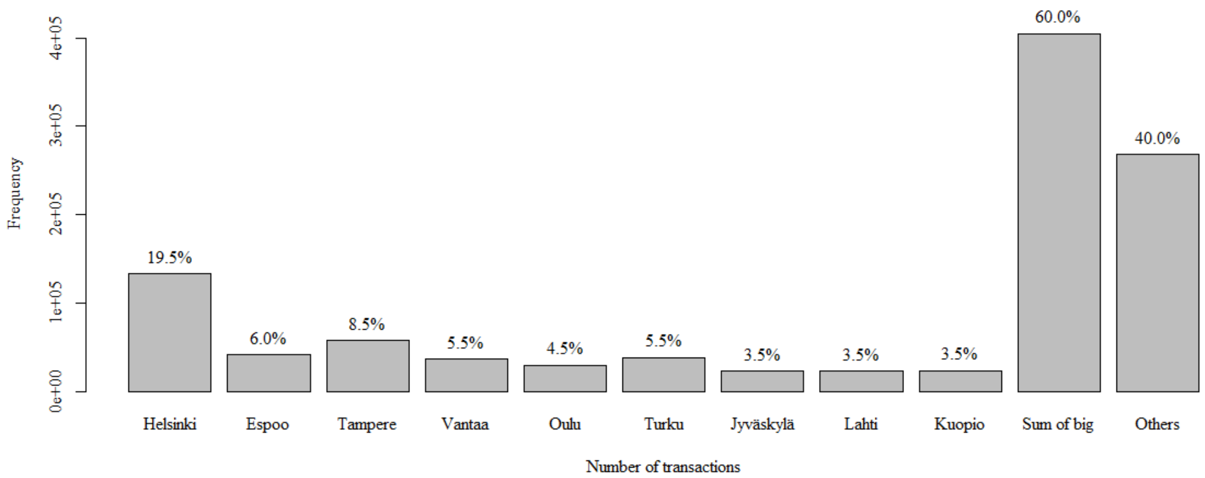
Figure 8 shows the distribution of transactions per year in our complete dataset. All other years are complete except for the year 2021 that only includes transactions for the first three months of the year.



Transactions are clustered in the larger municipalities as we can see from Figure 9. This is natural as most people in Finland live in those municipalities and thus, also a large portion of the transactions take place there. The nine largest municipalities account for 60 % of the whole transaction dataset we have.

Figure 9 – Distribution of the municipality

Figure 9 presents the areal distribution of the complete dataset. “Sum of big” consists of Helsinki, Espoo, Tampere, Vantaa, Oulu, Turku, Jyväskylä, Lahti and Kuopio. ”Others” includes the municipalities not belonging to the “Sum of big”.



4.2 Deriving the final dataset

We extracted the data from HSP with a total of 1 035 774 observations. The data represents actual realized dwelling transactions and their prices and characteristics, whereas the dwelling advertisement data (Oikotie.fi and Etuovi.com) would only present ask-prices of the dwellings, not actual realized ones. Generally speaking, the quality of the data is relatively high. Only professional real estate brokers and constructors get to add their dwelling transactions to the database, and we can assume that they are familiar with the housing related terminology. However, the data has been filled by hand, and consequently, the data includes some fallacious observations. In addition, the data includes rows that are missing some key variables, making these rows of data useless in our study. After removing all rows that have NAs in some relevant variable (variable used in the regressions) we have a total of 702 212 observations. These rows represent whole rows without any missing data but still include falsely inputted data by the real estate agents that we correct.

We start by identifying the variables that need adjustments. We use limit values to cut obvious mistakes from the lower and upper ends of each variable. In addition, we look in the data by hand. We examine all observations that the limits have not cut out but seem oddly large or small compared to other comparable apartments. If a value is deemed to being incorrect, we make adjustments to the data or remove the row depending on the case. After the data cleanup process, we are left with 673 362 observations that form our final dataset used in the regressions. Of the dataset, 501 217 observations consider dwellings in apartment buildings and 172 145 in row houses. Next, we go through every variable and explain the process of elimination and the corrections we make. Oikotie.fi and Etuovi.com dwelling advertisement sites are also used as a reference sample to check if certain observations are found currently in the market. With dummy variables we make sure that there is nothing but zeros or ones in the data. The variables that we examine are *Type, Price per square meter, Maintenance charge per square meter, Room number, Area, Age / Construction year, Total floors, Floor number, Bottom / Top floor*, *Condition* and *Sauna*. Other control variables are already in a usable form.

Type

Our dataset consists of two types of dwellings: the ones in apartment buildings and the ones in row houses. The data from KVKL's HSP represents the real estate agents' interpretations of the building types, but luckily the definitions of apartment buildings and row houses are well established. The interpretations of houses with balcony access (in Finnish *Luhtitalo*), low-rise apartment buildings (in Finnish *Pienkerrostalo*) and wooden houses (in Finnish *Puutalo*) however are not as clear. We remove all these housing types from the data because we are unable to decide where to allocate these rows (as apartment buildings or row houses) as in many cases these houses may be allocated to different categories depending on the definitions. Also, semi-detached houses (in Finnish *Paritalo*) are removed from the data as they are not fully comparable to row houses, as they have more yard space, only one common wall with the neighbor and better views (three directions instead of two), for example.

Price per square meter

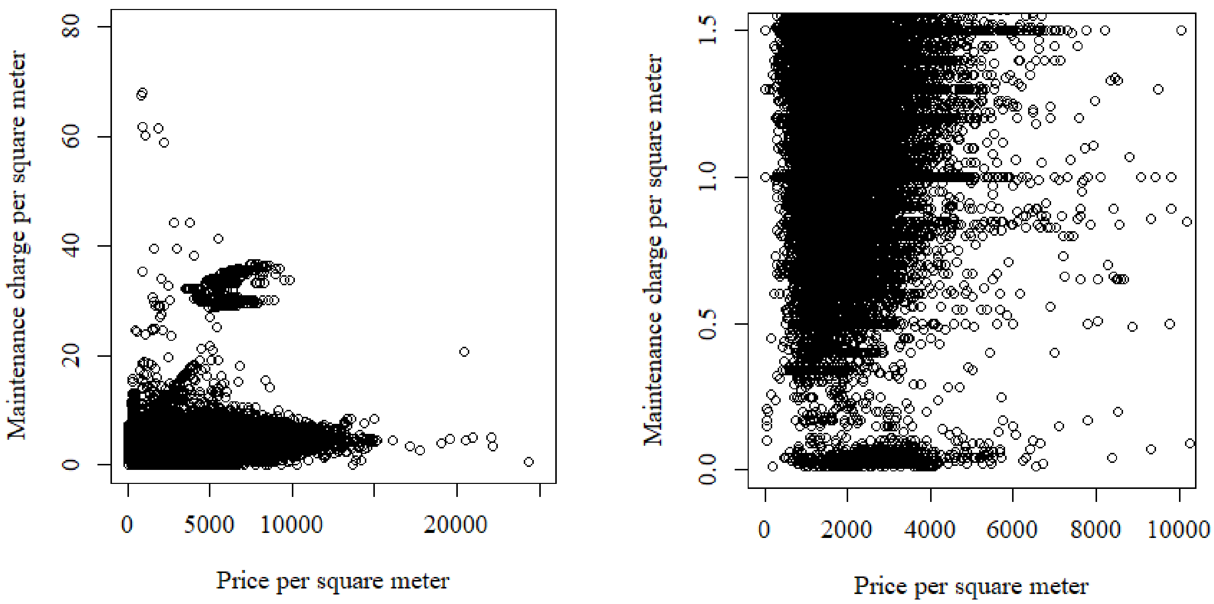
When we talk about *Price per square meter*, we refer to the debt-free price per square meter. With the lower end of the range of *Price per square meter* we have a few problems; many of these dwellings are right-of-occupancy dwellings (in Finnish *asumisoikeusasunto*) or dwellings that have characteristics that are not shown in our dataset. These extremely cheap dwellings that have characteristics that reduce the value of the dwellings we call "rotten apples". Such characteristics could include the obligation of destruction of the dwelling, for example. If these dwellings were included in the data sample, the regressions would suffer from omitted variable bias in the same way if the most expensive dwellings were included. For these two reasons we cut the lower end of the *price per square meter* variable out of the final data sample. We analyze the dwellings that are for sale at Etuovi.com at the moment and find that the highest price per square meter for a right-of occupancy dwelling in Finland is around 770e/sqm. We set the cutter for the lower end of *Price per square meter* at this point which reduces the sample by 26 608 dwellings. We assume that this cutter limits virtually all the right-of-occupancy dwellings as well as most of the rotten apples. By setting this cutter we are aware that some relevant data is lost, but the regression results stay unbiased, which is the main goal. At the upper end we set the cutter at 15 000 euros per square meter because of the same omitted variable bias. Eventually, these cutters together seem to enhance the regression statistics of our main regression models. We test if the model produces better estimates with a *Price per square meter* cutter set at 10 000 euros but find no difference to our 15 000-euro cutter. The residual plots from this test can be found from Appendix 5.

Maintenance charge per square meter

As we can see in Figure 10, there are some mysterious patterns in the *Maintenance charge per square meter* variable. In figure 10, both plots are from the same dataset, but the one on the right-hand side is just a zoomed version of the complete plot. The cluster around 33 euros per square meter in the maintenance charge is caused by two groups of new dwellings in two separate locations in the HMA. Both groups are built by the same constructor, and the maintenance charges on their website are significantly lower than the ones marked to the HSP database. Thus, those observations are deemed being incorrect and eventually removed. But we are still left with lots of outliers. It is, however, difficult to determine any hard cutter for the upper limit of the maintenance charge, as some dwellings are old ground level business spaces, that are often subject to doubled maintenance charges compared to the regular dwellings in the same housing company. Thus, even particularly high maintenance charges between 10 and 20 euros per square meters are possible. It is, however, difficult to imagine maintenance charges above 20 euros, and there are basically none listed on Etuovi.com at the moment. Thus, we find a cutter at 20-euro-level reasonable and decide to implement it. There is also a weird cluster below 0.1 euros per sqm (see Figure 10 on the right). We check around 20 random datapoints in the cluster by comparing them with existing dwelling advertisements in the same addresses on Etuovi.com and find that those are all clear mistakes that are caused by falsely marking the maintenance charge per square meter as the total maintenance charge. We remove these observations from the data. We test if the model produces better estimates with a maintenance charge per square meter cutter set at 10 euros but find no difference to our 20-euro cutter. The test residual plots can be found from Appendix 5.

Figure 10 – Price per square meter and Maintenance charge per square meter before cleaning the data

Figure 10 represents the relation between the *Price per square meter* and *Maintenance charge per square meter* including both apartment buildings and row houses before any data clean up. On the left is the whole dataset while on the right is the zoomed version of the same plot from the lower ends.

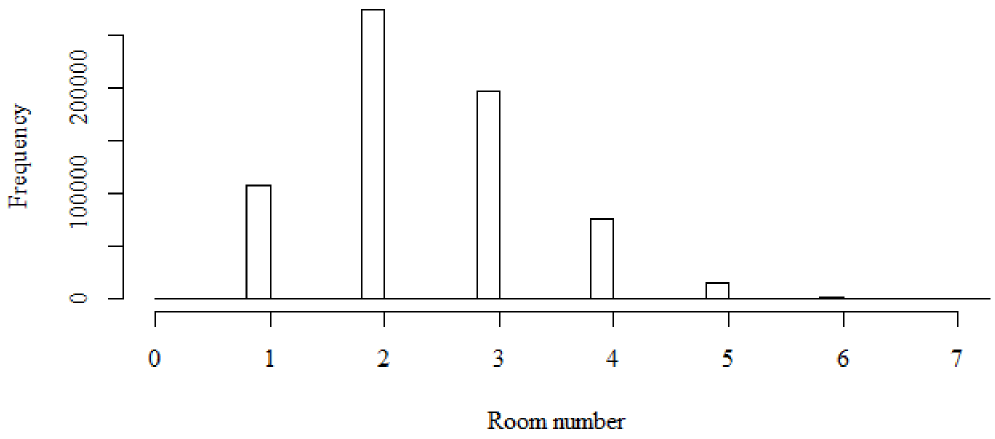


Room number

Extremely small and large rooms are exceptional. Figure 11 illustrates the distribution of the variable *Room number*. 99.6 % of the dwellings fall between one and five rooms. The lower limit for a room size is set by regulation. The Environment ministry has given a regulation *Suomen Rakentamismääräyskokoelma, Asuntosuunnittelu 1994 (G1)* de jure *construction law (rakennuslaki) 13 (557/89)*. According to the regulation, the minimum size for a room intended for living use is seven square meters. In our understanding this minimum size has been recognized by authorities before the 1994 regulation, but we could not find older documents from web sources. Also, it would not even make sense to construct a room being less than 7 square meters. Thus, we assume that all dwellings having an average room size less than seven square meters are typos and are removed from the dataset, but in addition to that, we also suspect that dwellings having an average room size between 7 and 15 square meters, or above 50 square meters might be incorrectly marked. All the 1864 dwellings falling into those categories (7 - 15 and above 50 square meter average room size) are looked in more detail and mistakes in the data are corrected.

Figure 11 – Distribution of the Room number after cleaning the data

Figure 11 shows the histogram of the variable *Room number* including dwellings in both apartment buildings and row houses. X-axis is capped at 7, but there are 91 dwellings having more than 7 rooms in the dataset outside of the scope of the histogram.



Area

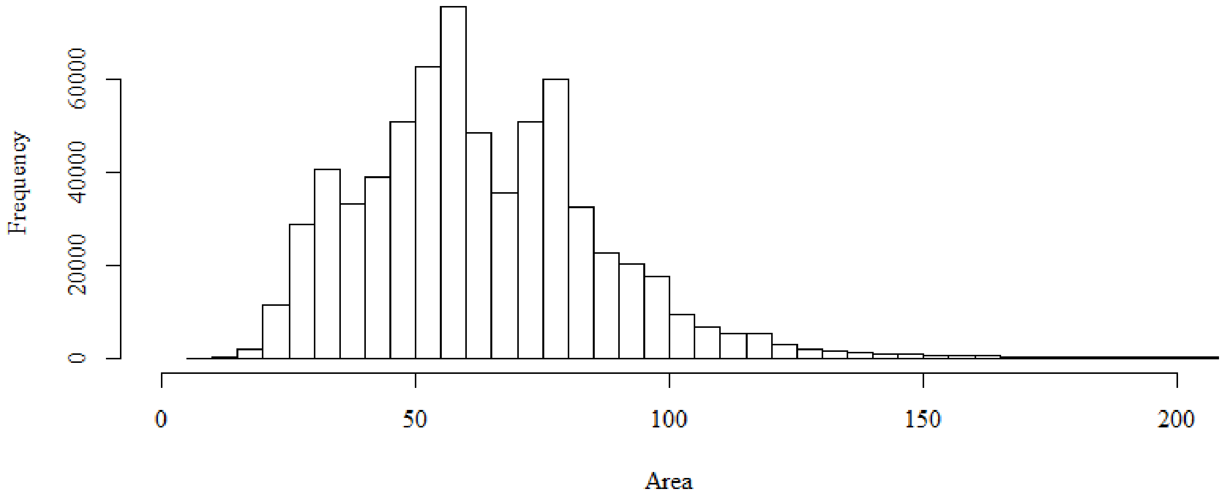
Particularly large living areas in our dataset often refer to transactions considering multiple dwellings sold all at once, or alternatively, they are typos. Our intention is to clear our dataset from both of those. By looking at the existing housing advertisements on Etuovi.com and Oikotie.fi, we find that dwellings in apartment buildings are very seldom over 280 square meters. Thus, we go through all the dwellings in apartment buildings having a living area of over 250 square meters, and over 300 square meters in row houses by hand. We compare the *Area* to the debt free price and the room description to find values that stick out from the crowd, examine them, and remove the mistakes.

The minimum room size in Finland is seven square meters, so we delete all datapoints having an area of less than 7 square meters. We set limits for the minimum size of the *area* according to the area of the smallest dwellings by room number listed on Etuovi.com. The following datapoints are removed:

- Single-room dwellings with less than 7 sqm
- 2-room dwellings with less than 26 sqm
- 3-room dwellings with less than 32 sqm
- 4-room dwellings with less than 60sqm
- 5-room dwellings with less than 82 sqm
- 6-room and bigger dwellings with less than 106 sqm

Figure 12 – Histogram of the Area after cleaning the data

Figure 12 shows a histogram of the *Area* after cleaning the data. Data includes dwellings in apartment buildings and row houses. X-axis is capped at 200 square meters for the sake of interpretability.

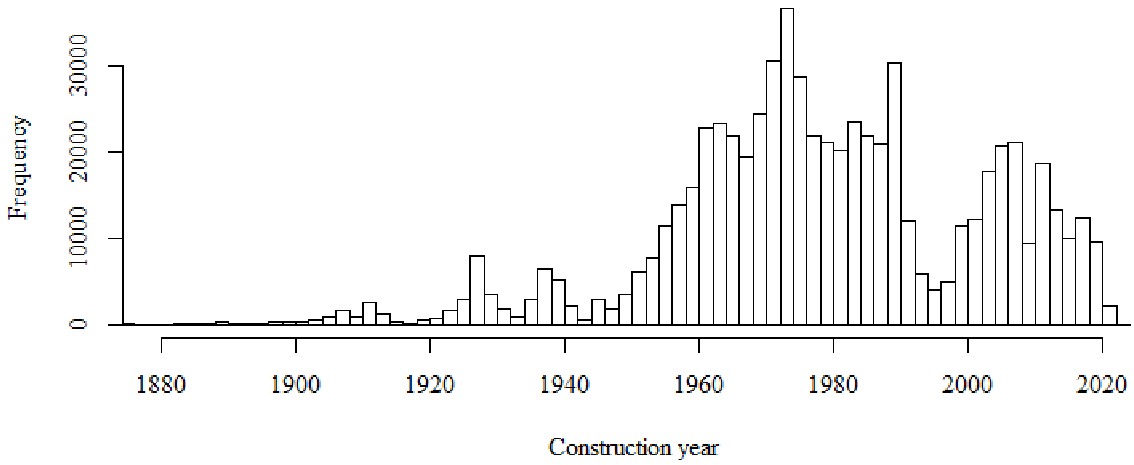


Age and Construction year

As the Finnish housing stock is, for the most part, built after the 1850's, we check all the 144 dwellings marked to have a construction year older than 1860 manually. We pay attention to the municipality in which the building is located as most old apartment buildings and row houses are located in a bunch of municipalities in reality (12 municipalities in our data sample). In borderline cases, we use Google Maps Street view to validate the construction year. Incorrect datapoints are being deleted in the process. We calculate the age of the dwelling by subtracting the construction year from the transaction year. With new built dwellings we get negative figures as many times new built dwellings are sold before the dwelling is finished. We assign 0 as age for new built dwellings that are sold before the completion of the construction process. If the construction year is empty, but the dwelling is new built, we use the transaction year as a construction year.

Figure 13 – Histogram of the construction year after cleaning the data

Figure 13 is the histogram of the *Construction year* after cleaning the data. Transactions considering dwellings in both, apartment buildings and row houses are included. The X-axis is capped at the lower end to year 1880 for the sake of interpretability. 354 dwellings in the final dataset have been constructed before year 1880.



Floor number & Total floors

Variable *Floor number* is only used with dwellings in apartment buildings, because row houses are always on the ground. In apartment buildings, the *Floor number* variable indicates the floor on which a dwelling is located.

Total floors variable, in turn, is used for different purposes with dwellings in apartment buildings and row houses. With apartment buildings, *Total floors* is used to refer to the total number of floors in the building, and it is only used in order to determine whether the dwelling is in a top floor (to calculate *Top floor dummy*). We believe that a second-floor dwelling is as valuable in a four-story building as it is in a building with eight floors, for example. Thus, the *Total floors* variable is not used as a control variable for apartment building regressions. Row houses are, however, sometimes multi-floor dwellings themselves, and thus the total number of floors may affect the value. For this reason, *Total floors* is used as a control variable in row house regressions.

We check all the dwellings in apartment buildings having a floor number or total floors in the building over 20 manually and delete incorrect datapoints. In order to validate some borderline cases, we check the addresses from the Google Maps Street View. With row houses, we remove all observations where total floors are reported being over three; we examine Etuovi.com and Oikotie.fi and find that none of even the most expensive dwellings in row houses have over three floors.

Top floor/Bottom floor

Top and *Bottom floor dummies* are used only for dwellings in apartment buildings, and they are formulated using the *Floor number* and *Total floors* variables. If *Floor number* is 1, the *Bottom floor dummy* gets value 1, and is 0 otherwise. If *Floor number* equals *Total floors*, the *Top floor dummy* gets value 1, and is 0 otherwise.

Condition

The real estate agents have estimated the condition of the dwellings using a 5-level scale. The original, verbal, descriptions are converted to number values as follows:

Bad = 1, Satisfactory = 2, Good = 3, Excellent = 4 and New = 5. Dwellings with unknown condition are omitted from the dataset.

Other dummy variables

The dataset includes ready-made dummy variables for *Balcony, Shore, Elevator, Sauna, Free-hold lot, Rented dwelling* and *New built dwelling*. We use these variables without alternations, except for the *Balcony* and *Sauna* variables. Balcony data contains both TRUE/FALSE and 1/0 values for the balcony value. We cannot know if the FALSEs are actually negative for *Balcony* or just missing values. In the data set, only 12 % of the dwellings are marked to have a balcony while on Etuovi.com, 65 % of the apartment building and row house dwellings have a balcony. We have to disregard the *Balcony* variable from the regressions. The case is more or less the same with the *Sauna* dummy, but we use the room description given in the data to find the dwellings with saunas and construct the dummies based on these descriptions. Balconies are not marked to the room description with as high consistency, so we cannot do the same to them. Every dummy variable gets value 1 when the dwelling has the attribute in question and value 0 when it does not have it.

Transaction time and locational variables

We use postal code and transaction year data as they appear in the data without making any adjustments. Our regression models use transaction time dummies on the level of years. Postal code level dummies are also included in every regression.

Sales time

If the sales time is missing in the dataset, we assign value 0 to the *Sales time* variable as we conclude that these sales are actually made in a zero-day lead time. Negative values as well as values over 1 000 days are eliminated, and these transactions are left outside of the regressions.

4.3 Descriptive statistics

Following Tables 2 and 3 present the descriptive statistics of the dataset separately for dwellings in apartment buildings and row houses.

Table 2 – Descriptive statistics of apartment building data – selected variables

Table 2 represents the descriptive statistics of apartment building data.

Statistic	N	Mean	St. Dev.	Min	Pctl(25)	Pctl(75)	Max
Price per square meter	501,217	2,666.514	1,563.680	770.000	1,510.420	3,369.570	15,000.000
Maintenance charge per square meter	501,217	3.230	1.008	0.100	2.500	3.800	19.960
Age	501,217	36.324	24.395	0	19	50	341
Area	501,217	57.643	21.934	7.000	42.500	71.000	452.500
Room number	501,217	2.236	0.916	0	2	3	15
Condition	501,217	2.745	0.690	1	2	3	5
Floor number	501,217	2.979	1.763	1	2	4	31
Total floors	501,217	4.877	2.090	3	3	6	35
Bottom floor	501,217	0.205	0.403	0	0	0	1
Top floor	501,217	0.244	0.430	0	0	0	1
Free hold plot	501,217	0.778	0.415	0	1	1	1
Rented dwelling	501,217	0.099	0.299	0	0	0	1
Elevator	501,217	0.533	0.499	0	0	1	1
Sauna	501,217	0.268	0.443	0	0	1	1
Shore	501,217	0.0002	0.014	0	0	0	1
New built dwelling	501,217	0.122	0.327	0	0	0	1

Table 3 – Descriptive statistics of row house data – selected variables

Table 3 represents the descriptive statistics of row house data.

Statistic	N	Mean	St. Dev.	Min	Pctl(25)	Pctl(75)	Max
Price per square meter	172,145	1,979.968	770.099	770	1,417.7	2,380.6	11,979
Maintenance charge per square meter	172,145	2.291	0.960	0.110	1.600	2.900	18.470
Age	172,145	21.792	14.308	0	11	31	300
Area	172,145	79.599	24.084	18	63	91	445
Room number	172,145	3.053	0.974	0	2	4	11
Condition	172,145	2.889	0.518	1	3	3	5
Floor number	172,145	1.000	0.000	1	1	1	1
Total floors	172,145	1.395	0.521	1	1	2	3
Bottom floor	172,145	0.000	0.000	0	0	0	0
Top floor	172,145	0.000	0.000	0	0	0	0
Free hold plot	172,145	0.755	0.430	0	1	1	1
Rented dwelling	172,145	0.056	0.229	0	0	0	1
Elevator	172,145	0.002	0.046	0	0	0	1
Sauna	172,145	0.831	0.375	0	1	1	1
Shore	172,145	0.001	0.026	0	0	0	1
New built dwelling	172,145	0.098	0.298	0	0	0	1

5 Methodology

5.1 OLS method and problems with the specification of the hedonic pricing model

To estimate hedonic pricing in the housing markets an ordinary least squares (OLS) regression is usually applied in the literature. Next, we dive deeper in the model specification and highlight some problems with the hedonic model that have been identified in previous literature. At the same time, we go through the basic requirements for the OLS model to work properly and assumptions that need to be made when using it.

The formulation of the model is one of the most important aspects of consideration with hedonic housing pricing studies. If the model specification is wrong, the model cannot explain the relation of the dependent and the independent variables, even if there is causation. As discussed earlier in section *3.1.4 – Hedonic pricing method*, Rosen (1976) and Lancaster (1966) can be considered the ones introducing the basic form of the hedonic pricing model. Researchers who have done comparisons to find superior model specifications compared to others have had little success. Dhrymes (1971, 96) suggests that the reason for not finding a superior form for the hedonic model in housing pricing context is the limited number of combinations across housing characteristics and spatial clustering of similar apartments. This can restrict the range of the sample so that only a small portion of the price surface is covered. In this case, many different forms of the model could have equal approximation power over that limited surface (Rothenberg, Galster, Butler & Pitkin 1991, 62). Many times, a semi-logarithmic (log-linear) form of the model is chosen as it seems to produce the best and easily interpretable results.

5.1.1 The linear relation between the dependent and independent variables

The OLS method is based on a few assumptions known as the Gauss-Markov assumptions. The first of the assumptions is that the dependent variable forms a linear relation with the explanatory variable(s). If the independent (explanatory) variables are non-linear one should transform them in a form that can be used in the OLS regression. Taking another functional form of the model, for example by using non-linear transformations of the independent variables, could help making the relation between the dependent and independent variables linear.

Sirmans et al. (2005) find that linear or semi-logarithmic models are usually used for housing pricing. Semi-logarithmic models typically use a log-form of the dependent variable but leave the independent variables unlogged. Follain and Malpezzi (1980) following the early studies on hedonic housing pricing develop the hedonic pricing model further. They test four functional forms for the model:

1) Logarithmic dependent variable with logarithmic independent variables;
2) Linear dependent variable with linear independent variables;
3) Logarithmic dependent variable with linear independent variables;
4) Linear independent variable with logarithmic independent variables;

of which 1) and 4) are disregarded due to poor usability with dummy variables. The use of semi-log models enables the researcher to use dummy variables. Follain and Malpezzi (1980) found little difference between models 2) and 3) in the explanatory power of these models but decided to use model 3) in their study for two reasons: The first one is that the explanatory power is slightly better than that of the fully linear model. Secondly, the log-linear specification also allows for variation in the dollar value of characteristics so that the additional value of a component is dependent on other characteristics of the house. (Follain and Malpezzi, 1980). Similarly, also Sirmans et al. (2005) state that the log-linear model allows for "variation in the characteristic prices across different price ranges within the sample" (Sirmans et al., 2005). Fox example, adding 10 square meters to a 20-square-meter dwelling should add more value than adding 10 square meters to a 200-square-meter dwelling. For variables that can be argued not being always linear, such as the age of the dwelling, Follain and Malpezzi (1980) test both linear-log and log-log models and find similar results.

A Box-Cox transformation of the variables is also possible. The Box-Cox transformation of the variables can be done if the effect of the independent variable on the dependent variable is not linear. The Box-Cox transformation is flexible, but highly sensitive to omitted variables. Cropper, Deck and McConnel (1988) study the use of Box-Cox transformation among other forms of the hedonic pricing model. They test six forms of the hedonic function and find that the linear and Box-Cox linear forms perform the best when there are missing variables present. The Box-Cox linear form is deemed to being the best and producing smallest average bias when estimating hedonic price functions. They conclude that "in general, when variables are omitted or replaced by proxies it is the simpler forms - the linear, semi-log, double-log and the Box-Cox linear that do the best" (Cropper et al., 1988). However, if a Box-Cox transformation is used, the interpretation of the coefficient estimates becomes more challenging. Even though the Box-Cox transformation usually explains the price variations well, the method may lead to imprecise estimates of the individual independent variables. We do not use the Box-Cox transformation for these reasons.

5.1.2 Random data generating process

One of the Gauss-Markov assumptions is that the observations come from a random process from the population. Our data contains 673 362 observations and represents the greater population very well. Sales made by individual sellers without the help of a real estate agent are not shown in the data and one could argue that dwellings sold by these sellers have different characteristic combinations and/or are priced differently when there are no professional parties included. However, these transactions form a relatively small group of observations compared to the whole population as most dwelling transactions are conducted with the help of real estate agents.

5.1.3 Normality assumption of the error terms

The next assumption is that the error terms are normally distributed. If the error terms are not normally distributed it makes the coefficient estimates biased. Also, the distribution of which the error terms form should have a mean of 0. This means that there is no relationship with the independent variables and the errors terms. When the error terms are random and have no relation with the independent variables the coefficient estimates are unbiased.

Often, outliers in the data make the error terms not normally distributed. An easy way to deal with such a problem is to remove these outliers from the data. If observations are removed from the data, it should be done with caution as some of the information value of the regression is also lost. Especially, if the dependent side values are outliers in the data, removing such observations could be a very poor decision. It is very possible that the model just does not include all relevant independent variables that would explain the dependent variables' values that seem exceptional. However, if those omitted variables cannot be included in the model, the removal of the outliers is justified.

The situation where variables outside of the scope of the model are correlated with a variable included in the model leads to an endogeneity problem. When omitted variables are correlating with independent variables the effect will be seen in the error term. The assumption of no correlation between the error term and the independent variables does not hold and the OLS model generates biased estimates of the regression coefficients. The independent variables pick up the effect the model cannot assign on the omitted variable.

Our data is limited in such a way that adding independent variables outside of HSP is quite hard. Structural characteristics that are not included in the data from HSP are virtually impossible to add in the data. One of the biggest flaws in our data is that there is no information of the renovation backlog (in Finnish *Korjausvelka*), and some unusual dependent values could be explained by adding the renovation backlog in the model. By renovation backlog, we mean the depreciation of the dwelling, especially the water/sewer pipe depreciation and other large-scale depreciations that will lead to large scale renovations in the future. If the pipes are repaired 50 years ago, then the dwelling has lots of renovation backlog as the pipe repair is again just around the corner and should affect the value of the dwelling significantly.

5.1.4 Heteroscedasticity problem in the hedonic model

In addition to random normally distributed error terms it is expected that the error terms' variance is constant across observations. Heteroscedasticity occurs when the error terms of the model do not have a constant variance; each residual term comes from a different distribution and that violates the normal distribution assumption of the error terms in the model (Stevenson, 2004). There should also be no correlation with different error terms (no autocorrelation). If there is no heteroscedasticity nor autocorrelation, the error terms are independent and identically distributed (IID) and the regression model produces efficient coefficient estimates, even if they are biased. It is essential that we define the functional form of the model in such a way that it minimizes the heteroscedasticity problem.

Using a semi-logarithmic model helps with a heteroscedasticity problem that hedonic housing pricing models often suffer from. The effects of the independent variables to the price are not the same for cheap and expensive dwellings; when the dwelling price in the data sample increases there are often factors affecting the prices that are not included in the model resulting in heteroscedasticity. Also missing variables and measurement errors in the data make the model sensitive for heteroscedasticity.

Autocorrelation is often a problem with time series and cross-sectional data. In a hedonic housing pricing context spatial autocorrelation often occurs as a spatially correlated independent variable is omitted from the model. For example, dwellings located in the central Helsinki are particularly expensive due to the central location, and structural characteristics of these dwellings cannot explain the high prices alone. Spatially clustered omitted variables produce spatially clustered error terms. This in turn makes the variance estimates of the independent variables biased, usually making them too small. As a result, the t-test statistic shows higher statistical significance than it should. In this case neighborhood level dummy variables taking in account these fixed effects are often introduced in the model as they alleviate the problem (Kuminoff, Parmeter & Pope, 2010). We use a fixed effects model with postal codes (basically postal code dummies) to deal with this problem. We also cluster the standard errors on the postal code and year level to avoid possible problems with heteroscedasticity.

The heteroscedasticity problem is closely linked to the concept of submarkets introduced in section *4.1.5 – Regional submarkets*. Error terms may be clustered within district-specific locational submarkets, such as postal codes, but also within other kinds of submarkets. Houses with the view of the sea, for instance, may form a submarket from this perspective. Even though if the dwellings are located in the same postal code area, they do not necessarily belong to the same submarket form this perspective as they may not be close substitutes for each other. But as discussed, recognizing the subgroups is particularly difficult. If the subgroups were correctly distinguished, there would be no heteroscedasticity problem as the subgroup dummies would capture the effect that cannot be assigned to other independent variables.

5.1.5 Multicollinearity

The OLS model cannot have perfect multicollinearity between variables. Multicollinearity means that independent variables correlate with each other. With multiple OLS regressions where there are many independent variables this problem sometimes occurs. One should add only variables that have a clear intuitive reason to affect the dependent variable. Correlating variables make the model inefficient; the coefficient of determination decreases as the standard errors of the model increase. If there is a clear correlation between variables, one should consider removing the correlating variable from the regression. Taking the correlating variable out is problematic, however. As discussed above in section *5.1.3 – Normality assumption of the error terms*, omitting a correlating variable produces biased coefficient estimates. Also, with multicollinearity the coefficient estimates become rather sensitive to changes in the data. Often correlations close to 90 % are deemed being too high and perfect multicollinearity, meaning that independent variables correlate perfectly, cannot even be included in an OLS regression. In case of perfect multicollinearity, the OLS method cannot determine what is the effect of the individual independent variables on the dependent variable.

5.2 Determining the form of the variables

Price

The price of a dwelling is the dependent variable in the model. As discussed in the section *5.1 – OLS method and problems with the specification of the hedonic pricing model*, the main question regarding the price variable is the specification of the form; should we use logged or unlogged form? Also, we must choose whether to use price per square meter or plain price. We choose the price per square meter because it offers us more easily interpretable and generalizable results. Also our variable of interest, the maintenance charge, is divided by the area and thus their relation is easy to understand.

Then we must decide whether to use the logarithmic form of the variable *Price per square meter*. The OLS assumptions are discussed in detail in section *5.1. – OLS method and problems with the specification of the hedonic pricing model*. Of these six assumptions, the assumptions of linear relation between dependent and independent variable, homoscedasticity and normality of the error terms can be analyzed by plotting the regression residuals in different ways. We use standardized residuals in order to maintain the comparability between the models. The standardized residuals vs fitted values plots tell us whether the relation of the independent and dependent variables is linear in form, and whether the residuals have constant variance across the sample. The residual vs fitted values plots and the detailed discussion related to them are show in Appendix 2. The Q-Q plots tell whether the residuals are normally distributed. The Q-Q plots and their interpretations can be found from Appendix 3. In addition, distribution plots of the standardized residuals tell the same story as the Q-Q plots and can be found in Appendix 4 with the related discussion. To summarize the results of the residual analysis, we decide to use the logarithmic transformation of the dependent variable in our regressions as it fulfils the normality and homoscedasticity assumptions of the error terms better than the unlogged form. The logarithmic transformation also fulfils the assumption of linear relation between dependent and the independent variables better than the unlogged version.

Maintenance charge per square meter

The costs of a housing company are distributed to the individual dwellings according to its share of the shares of the whole housing company. The number of shares is usually in relation to the area of the dwelling. Thus, using maintenance charge per square meter is easier to use than plain maintenance charge as the common costs are principally distributed according to the relative dwelling size to each dwelling. Using per square meter values of the maintenance charge and price of the dwelling makes comparisons between different levels of maintenance charges and dwelling prices easier. In our main regressions we are interested in how much changes in the maintenance charge per square meter affect the price per square meter on average. In further analysis we also test the model with maintenance charge per square meter dummy variables as part of interaction terms in order to find whether capitalization of the maintenance charges is linear or not across the spectrum of maintenance charges. A log transformation of the variable *Maintenance charge per square meter* would tell us what the effect of a one percent change in the maintenance charge is on the dwelling price on average. We are only interested in the euro effects and do not use it.

Age

One of the most problematic variables is the *Age* of the dwellings for a couple of reasons that also happen to be connected with each other. The first problem is the non-monotonicity & non-linearity of the variable, and the second is heteroscedasticity. The non-linear nature of *Age* means that the marginal effect of time on *Age* decreases as the dwelling gets older. The non-monotonicity problem is even trickier. When the dwellings ages it starts to depreciate at first but passing a critical point it actually starts to appreciate in value the older it gets. This non-monotonic nature of the *Age* is what Goodman and Thibodeau (1995) call the vintage effect. In addition to the vintage effect, housing company level renovations, such as pipe repairs, increase the value of the individual dwellings significantly once they are made also resulting in a non-monotonous behavior of the *Age* variable. Consequently, also Laakso (1997) finds a U-shaped relation of age and price of a dwelling in the HMA.

Thus, the variety in housing company level renovations is closely linked to the variable A*ge*, and according to Goodman and Thibodeau (1995, 1997), the *Age* variable is the primary cause of heteroscedasticity in hedonic housing pricing models. Also, according to Wilhemsson (2008) the lack of precise information of the renovations causes heteroscedasticity in the models. We suffer from the lack of this renovation data, but as Stevenson (2004) puts it simply, the data of housing company level renovations is seldom available nor used in research.

With the variable *Age* we have basically four alternative forms to use it in the regression models: plain *Age*, log *Age*, power transformations of *Age* and *Age* dummies in a few different ways. Plain *Age* tackles the problems particularly poorly. The non-linearity & non-monotonicity problems could be handled using power transformations, but second and third powers of *Age* show high correlations (around 95 percent) together with the first power which is a severe problem as discussed in section *5.1.5 – Multicollinearity*. Log transformation tackles the non-linearity problem particularly well. However, it does not take into account the non-monotonicity, nor it solves the heteroscedasticity problem. The last option, *Age* dummies take in account the non-linearity and non-monotonicity problems, and we find them the best solution to the heteroscedasticity problem as well. As major housing company level renovations are usually executed in the same age across dwellings, *Age* dummies might capture the effect of these renovations at least to some extent. We decide to test the regression models with both, log *Age* and *Age* dummies, and find that the regression residuals meet the OLS requirements equally well. Since we find the arguments for *Age* dummies most reasonable, we decide to go with them instead of the log transformation. We test different splits of the *Age* dummies to see if a more frequent or sparse division of the dummies fits the data better but observe virtually no difference. A denser division of the dummies seems more reasonable from a practical standpoint, as often major renovations are executed in the same age across all dwellings.

We set the *Age* dummies as follows:

0-5, 6-10, 11-15, 16-25, 26-35, 36-45, 46-55, 56-65, 66-80, 81-95, and over 95 years.

Area

Laakso (1997) finds that the price of a dwelling increases the bigger the dwelling is, and that the price per square meter decreases with size. There are similar problems arising with the *Area* as there are with *Age*. The relationship of the *Area* and the *Price per square meter* may not be linear. We test both log-form and unlogged form of the variable *Area*. In addition, second and third powers of the *Area* were tested alongside with the first power, but strong correlations (above 95 %) between these variables are observed and we decide not to use them in our models. Also Laakso (1997) finds that with a semi-log model, like ours, using second and third powers of the *Area* is clearly a wrong functional specification. Thus, we decide to go with the log version of *Area*.

Other control variables

We test our model with room number dummies ranging from 1 to 8+ rooms and find that 6, 7 and 8 rooms have no statistically significant explanatory power. We decide to continue with the *Room number* dummy variables but use only six dummies (1 - 6+ rooms). The condition data is originally on a five-step scale, but we assume that the effect is not linear throughout the scale. We tackle the problem by forming *Condition* dummies. As we assume that the marginal price effect of moving one floor up or down in an apartment building (regardless of the floor the dwelling is on) is constant, we use the *Floor number* as a continuous control variable in the regression models regardless of the discrete form of the data. As discussed in section *4.2 – Deriving the final dataset*, the *Floor number* is only used with apartment building data. *Total floors* is similarly used as a continuous variable, but only for row house data. *Top* and *bottom floor* dummy variables are used in the apartment building models as intuitively it seems reasonable to assume that penthouses are priced higher and ground levels lower than otherwise comparable dwellings taking in account the floor number. *Elevator, Sauna* and *Shore* dummy variables are also included. *Shore* does not mean having own shore but instead being in the near presence of a shore. However, the notation of *Shore* on HSP is a bit vague as any clear distance thresholds are not set. *Free-hold plot* dummy is included in the model even though some of the effect of *Free-hold plot* is included already in the *Maintenance charge per square meter* variable. However, they do not seem to correlate adversely. We also employ a *New built dwelling* dummy since dwellings with age of zero years can be brand new or used depending on the case; it takes one year until the age turns from zero to one. *Rented dwelling* dummy is included as buyers looking for a dwelling for themselves probably prefer dwellings that are not rented. Otherwise, they would need to wait for the end of the term of notice before getting to move in.

Location

One of the most important factors in our regression model is location. In section *3.1.5 – Regional submarkets* we discussed the topic of submarkets and concluded that the market consists of small submarkets within which dwellings are close substitutes for each other. Location is one and probably the most obvious way to form submarkets and is usually done by postal code level in the literature. One could argue that postal codes separate different neighborhood and locational characteristics into subgroups. However, differentiating locational and neighborhood characteristics in different subgroups with postal codes is not a perfect solution. Inside these postal codes very different micro-locations can be found with different mixture of characteristics. The postal code areas are just lines drawn to the map quite randomly. We understand the limitations of our postal code fixed effects but use this approach due to its simplicity.

Time

In order to catch the systematic changes in price levels over time that are caused by, for example macroeconomics, we employ time fixed effects at year level.

5.3 Correlations between the variables

We test the correlations between the variables in order to avoid the multicollinearity problem. We do not find alarmingly high correlations. Highest correlations exist between *Room number 1* and

Ln(Area) (-0.7), *Condition good* and *Condition satisfactory* dummies (-0.85), *New build dwelling* and *Age 0 to 5* (0.8), and finally between *Elevator* and *Total floors* (0.66). The OLS method cannot be used with perfect multicollinearity, but correlations around 80 - 90 % are often observed with hedonic housing pricing models in the literature. Thus, we conclude that our model can be used keeping in mind that there are some relatively high correlations involved, which may lead to inefficient coefficient estimates. The correlation matrix of the final variables in our main models for the combined dataset (apartment buildings and row houses) can be found in Appendix 1.

6 Results

We test our research questions in this section. The main regression results are not particularly consistent with previous research, as we find higher discount rates associated with the maintenance charge per square meter compared to rates that are typically found with operational costs. In other words, we find dwellings overvalued w.r.t maintenance charges. We start by presenting our regression formula and how to calculate the effects implied by the regression coefficients on the dwelling prices. Then we proceed to the main regression and finally cover some interesting additional analyses.

6.1 Main regression formula

The form of the regression variables has been covered in section *5.2 – Determining the form of the variables*. Putting all the variables together, the main model specification is for apartment building data as follows:

$$
\begin{aligned}
&Ln(Price\ per\ square\ meter) \\
&\quad = \beta_1 * \ Maintenance\ charge\ per\ sqm + \beta_2 * ln(Area) + \beta_3 \\
&\quad * Room\ number\ 2 + \beta_4 * Room\ number\ 3 + \beta_5 \\
&\quad * Room\ number\ 4 + \beta_6 * Room\ number\ 5 + \beta_7 \\
&\quad * Room\ number\ 6\ or\ more + \beta_8 * Age\ 6\ to\ 10 + \beta_9 \\
&\quad * Age\ 11\ to\ 15 + \beta_{10} * Age\ 16\ to\ 25 + \beta_{11} * Age\ 26\ to\ 35 + \beta_{12} \\
&\quad * Age\ 36\ to\ 45 + \beta_{13} * Age\ 56\ to\ 55 + \beta_{14} * Age\ 56\ to\ 65 + \beta_{15} \\
&\quad * Age\ 66\ to\ 80 + \beta_{16} * Age\ 81\ to\ 95 + \beta_{17} * Age\ over\ 95 + \beta_{18} \\
&\quad * Condition\ satisfactory + \beta_{19} * Condition\ good + \beta_{20} \\
&\quad * Condition\ excellent + \beta_{21} * Condition\ new + \beta_{22} * Sauna \\
&\quad + \beta_{23} * Free\ hold\ plot + \beta_{24} * New\ built\ dwelling + \beta_{25} \\
&\quad * Shore + \beta_{26} * Rented\ dwelling + \beta_{27} * \boldsymbol{Floor\ number} + \beta_{28} \\
&\quad * \boldsymbol{Bottom\ floor} + \beta_{29} * \boldsymbol{Top\ floor} + \beta_{30} * \boldsymbol{Elevator} \\
&\quad + \ Year\ fixed\ effects + Postal\ code\ fixed\ effects + \ \varepsilon
\end{aligned}
\tag{2}
$$

And for row house data as follows:

$$
\begin{aligned}
&Ln(Price\ per\ square\ meter) \\
&\quad = \beta_1 * Maintenance\ charge\ per\ sqm + \beta_2 * \ln(\text{Area}) + \beta_3 \\
&\quad * Room\ number\ 2 + \beta_4 * Room\ number\ 3 + \beta_5 \\
&\quad * Room\ number\ 4 + \beta_6 * Room\ number\ 5 + \beta_7 \\
&\quad * Room\ number\ 6\ or\ more + \beta_8 * Age\ 6\ to\ 10 + \beta_9 \\
&\quad * Age\ 11\ to\ 15 + \beta_{10} * Age\ 16\ to\ 25 + \beta_{11} * Age\ 26\ to\ 35 + \beta_{12} \\
&\quad * Age\ 36\ to\ 45 + \beta_{13} * Age\ 56\ to\ 55 + \beta_{14} * Age\ 56\ to\ 65 + \beta_{15} \\
&\quad * Age\ 66\ to\ 80 + \beta_{16} * Age\ 81\ to\ 95 + \beta_{17} * Age\ over\ 95 + \beta_{18} \\
&\quad * Condition\ satisfactory + \beta_{19} * Condition\ good + \beta_{20} \\
&\quad * Condition\ excellent + \beta_{21} * Condition\ new + \beta_{22} * Sauna \\
&\quad + \beta_{23} * Free\ hold\ plot + \beta_{24} * New\ built\ dwelling + \beta_{25} \\
&\quad * Shore + \beta_{26} * Rented\ dwelling + \beta_{27} * \boldsymbol{Total\ floors} \\
&\quad + Year\ fixed\ effects + Postal\ code\ fixed\ effects + \varepsilon
\end{aligned} \tag{3}
$$

The bolded variables highlight the differences between the equations (2) and (3).

6.2 Coefficient transformations

When using a logged dependent variable, the interpretation of the independent variable coefficient is quite straight forward but requires that we first transform them into precise form; the coefficients tell only the effect on the logged dependent variable, but we want to know the effect on the dependent variable itself. The transformation is done differently to logged and unlogged independent variable coefficients. To transform unlogged independent variable coefficients to interpretable form we need to exponentiate them to get precise values, instead of approximations. Below is the formula for exponentiation:

$$
Coefficient_{Unlogged-Transformed} = (e^{Coefficient_{unlogged}} - 1) * 100\ \% \tag{4}
$$

To change logged independent variable coefficients into interpretable and precise form we need to do another transformation. We want to know the effect of a one percent change in the independent variable on the dependent variable. To do this we use the below formula:

$$Coefficient_{Logged-Transformed} = (1.01^{Coefficient_{logged}} - 1) * 100\ \% \qquad (5)$$

Now the transformed coefficients' interpretations are:

$Coefficient_{Unlogged-Transformed}$: A one unit increase in the independent variable has a $Coefficient_{Unlogged-Transformed}$ percentage effect on the geometric mean of the dependent variable *Price per square meter*.

$Coefficient_{Unlogged-Transformed}$: A one percent change in the independent variable has a $Coefficient_{Logged-Transformed}$ percentage effect on the geometric mean of dependent variable *Price per square meter*.

6.3 The observed effect of maintenance charges on dwelling prices

We run the main regressions separately with apartment building data (models (1) and (3)) and row house data (models (2) and (4)). Models (1) and (2) use the *Price per square meter* as the dependent variable. Our main regression models (3) and (4) follow the equations (2) and (3), respectively, and take a natural logarithm of the *Price per square meter*. Models (1) and (2) are presented only for the sake of showing how poorly the unlogged dependent variable works. The model (1) maintenance charge coefficient is positive and when using the correct model (3) it turns correctly to negative. They have same independent variables as models (3) and (4). From now on when talking about the model results, we refer only to models (3) and (4). Models (1) and (2) are not discussed any further. We do the coefficient transformation presented in section *6.2 – Coefficient transformations* automatically, thus with large unlogged coefficient values (above 0.1) the transformed value starts to deviate from what the unlogged coefficient shows. With the maintenance charge especially, we want to remind the reader that it is a monthly charge before going in the results. Table 4 summarizes the results from the four regression models.

Table 4 - Effect of maintenance charges on dwelling prices

Table 4 represents the results of our main regression models. Models (1) and (3) use the apartment building data while models (2) and (4) use the row house data. Models (1) and (2) have plain *Price per square meter* as a dependent variable whereas in models (3) and (4) we take a natural logarithm of the *Price per square meter*. All models have postal code and year level fixed effect dummies included. Also, standard errors are clustered at postal code and year level. Standard errors are show in parentheses below the coefficient estimates. ***, ** and * note statistical significance at 1 %, 5 % and 10 % levels, respectively.

	Dependent variable:			
	Price per sqm		Ln(Price per sqm)	
	(1)	(2)	(3)	(4)
Data	Apartment building	Row house	Apartment building	Row house
Maintenance charge per sqm	0.819	-15.235***	-0.012***	-0.012***
	(10.543)	(3.980)	(0.002)	(0.002)
Ln(Area)	-950.708***	-630.521***	-0.319***	-0.290***
	(96.874)	(42.738)	(0.017)	(0.010)
Room number 1	ref.	ref.	ref.	ref.
Room number 2	38.902	126.044***	-0.003	0.073***
	(24.693)	(15.940)	(0.005)	(0.007)
Room number 3	208.978***	244.502***	0.037***	0.128***
	(43.268)	(21.633)	(0.008)	(0.009)
Room number 4	371.394***	282.717***	0.083***	0.145***
	(56.253)	(26.056)	(0.009)	(0.010)
Room number 5	545.558***	299.656***	0.157***	0.154***
	(63.614)	(28.384)	(0.013)	(0.011)
Room number 6 or more	717.793***	344.081***	0.229***	0.180***
	(100.602)	(33.059)	(0.022)	(0.014)
Age 0 to 5	ref.	ref.	ref.	ref.
Age 6 to 10	-227.368***	-156.747***	-0.077***	-0.070***
	(25.412)	(17.072)	(0.005)	(0.005)
Age 11 to 15	-433.677***	-294.574***	-0.156***	-0.140***
	(31.489)	(22.518)	(0.012)	(0.006)
Age 16 to 25	-671.260***	-428.421***	-0.253***	-0.211***
	(39.684)	(27.036)	(0.010)	(0.007)
Age 26 to 35	-928.359***	-577.303***	-0.365***	-0.291***
	(53.266)	(33.769)	(0.011)	(0.008)
Age 36 to 45	-1,206.660***	-752.976***	-0.441***	-0.379***
	(75.572)	(43.965)	(0.019)	(0.014)
Age 46 to 55	-1,340.570***	-816.289***	-0.456***	-0.410***
	(98.366)	(64.875)	(0.023)	(0.024)
Age 56 to 65	-1,130.321***	-575.635***	-0.388***	-0.330***

	(78.179)	(74.605)	(0.017)	(0.023)
Age 66 to80	-1,202.487***	-731.003***	-0.385***	-0.384***
	(112.286)	(67.995)	(0.021)	(0.024)
Age 81 to 95	-657.990***	-539.643***	-0.282***	-0.277***
	(107.237)	(55.423)	(0.020)	(0.025)
Age over 95	-396.078***	-231.104**	-0.219***	-0.153***
	(79.531)	(110.617)	(0.019)	(0.047)
Floor number	56.037***		0.015***	
	(5.736)		(0.001)	
Bottom floor	28.516***		0.003*	
	(6.092)		(0.001)	
Top floor	-19.791**		-0.007***	
	(8.041)		(0.002)	
Total floors		-42.980***		-0.024***
		(4.870)		(0.002)
Condition bad	ref.	ref.	ref.	ref.
Condition satisfactory	294.669***	260.458***	0.085***	0.139***
	(30.283)	(22.208)	(0.004)	(0.008)
Condition good	550.662***	447.995***	0.190***	0.242***
	(46.940)	(35.900)	(0.007)	(0.013)
Condition excellent	910.316***	670.333***	0.259***	0.315***
	(84.677)	(35.461)	(0.018)	(0.013)
Condition new	967.102***	727.083***	0.264***	0.298***
	(62.095)	(38.932)	(0.014)	(0.013)
Elevator	-5.677		0.024***	
	(19.144)		(0.005)	
Sauna	38.610*	1.642	0.068***	0.009***
	(21.462)	(6.494)	(0.005)	(0.003)
Shore	19.288	442.456***	0.092***	0.180***
	(125.024)	(59.560)	(0.029)	(0.023)
Free hold plot	181.527***	90.830***	0.058***	0.044***
	(29.008)	(10.262)	(0.007)	(0.004)
New built dwelling	101.064*	82.448**	0.044***	0.051***
	(49.251)	(30.601)	(0.012)	(0.008)
Rented dwelling	-19.508	-17.289**	-0.008***	-0.009**
	(11.626)	(6.590)	(0.003)	(0.004)
Year dummies	Yes	Yes	Yes	Yes
Postal code dummies	Yes	Yes	Yes	Yes
Observations	501,217	172,145	501,217	172,145
R^2	0.860	0.867	0.907	0.887
Adjusted R^2	0.860	0.866	0.907	0.886

Residual Std. Error	584.996 (df = 500109)	281.575 (df = 170745)	0.167 (df = 500109)	0.127 (df = 170745)
Note:				*p**p***p<0.01

As the table shows, the coefficient estimates are statistically very significant on general level. *Maintenance charge per square meter* coefficients get values -0.012 (***) in both models (3) and (4). The maintenance charge per square meter has a clear, but small impact on dwelling prices; the estimates suggest that a one euro increase in the monthly maintenance charge per square meter decreases the price per square meter by -1.2 % on average with both dwellings in apartment buildings and row houses. The geometric mean of dwelling price per square meter for dwellings in apartment buildings in our data sample is 2292.2 euros, meaning that on average, a one euro increase in the maintenance charge per square meter decreases the dwelling price per square meter by approximately 28 euros. The geometric mean of price per square meter for dwellings in row houses is 1844.3 euros and thus a one euro increase in the maintenance charge per square meter decreases the dwelling price per square meter by approximately 22 euros with row houses.

We use the following formula to calculate the implied discount rates of the maintenance charges:

$$IDL = \frac{One\ euro\ change\ in\ the\ mothly\ maint.charge\ per\ sqm * 12\ Mon.}{Geometric\ mean\ dwelling\ price\ per\ sqm *\ Coefficient_{Unlogged-Transformed}} \qquad (6)$$

In which $Coefficient_{Unlogged-Transformed}$ is the regression coefficient transformed with equation (4), and *IDL* represents the implied discount rate.

The numbers above (28 euros and 22 euros) turn into 43 % and 55 % yearly implied discount rates for dwellings in apartment buildings and row houses, respectively, when discounting to perpetuity with the formula presented above. The discount rates are extremely high in the light of financial theory, and also very high compared to previous literature on operational housing costs (see e.g. Longstreth et al. (1985), Dinan & Miranowski (1986), Janssen (2003) and Kahn & Kok (2014)). However, one must remember that we use an infinite discount period, which naturally increases the implied discount rate compared to shorter periods.

When comparing the implied discount rates to the typical required rates of return of stock and bond investments observed in the market, we find that our implied discount rates are substantially higher. Traditionally stocks have been considered as rather risky investments, but still the required rates of return are typically between 5 – 10 %. Before comparing our implied discount rates to these numbers, we must have an understanding of the "riskiness" (stability) of the maintenance charges. Figures 6 and 7 in section *2.1.2 – Housing company charges* show how the expenses and income of housing companies are typically constructed. The stability of the maintenance charges is derived from the stability of the costs of the housing company and the income of the housing company (not including the maintenance charges) as it is the result of their difference. It is relatively safe to assume that none of the largest components that construct the income (excluding the maintenance charges), nor the expenses fluctuate in large amounts over the years. For example, the heating costs are quite stabile over time, even if some years might be colder/hotter than others. Another example is the plot leases, that are in fact stabile over a long period of time (e.g. for decades). In addition, if we think of the rental income of the housing companies for example, the required rate of return from such investment should be the required rate of return used in real estate investments, which is also rather low. Now, if the expenses are quite stabile and also the income is rather stabile, so should be the maintenance charges as it is the plug in this equation. Hence, taking in account the "riskiness" of the level and stability of maintenance charges, the implied discount rates we find are especially high.

The reason for observing such high implied discount rates remains unclear in the main regressions. We hypothesize that the effect may not be constant across Finland. In the most densely populated areas where housing demand overcomes supply and buyers need to be fast in their actions, factors such as the maintenance charges may not be among the most important characteristics under consideration when buying a dwelling. Later in the section *6.5 – Locational differences in the effect of maintenance charges* we see that the observed small effect of -1.2 % is actually derived from the largest municipalities in Finland.

Other control variables

The other control variables receive coefficients that are relatively in line with our expectations. *Ln(Area)* coefficients get values -0.391(***) in model (3) and -0.290 (***) in model (4). The coefficients represent the elasticity of the *Price per square meter* with respect to *Area*. One percent increases in the living area of dwellings have a -0.39 % (apartment buildings) and a -0.29 % (row houses) decreasing effect on the price per square meter.

As a whole, the *Room number* coefficients increase when additional rooms are added, as one would expect. We control for the area meaning that the trend of increasing price per square meter when adding additional rooms is independent of the total area. The explanation for observing such behavior might be better usage of space; if you can put four rooms in a 70 square meter dwelling instead of just three, it creates more opportunities for different buyers to consider the dwelling. For example, families with two kids would prefer a four-room dwelling over three rooms. In addition, every room often brings an additional window that could also explain the observed results.

The *Age* coefficients show expected patterns. The dwelling value decreases when the dwelling ages up until the age of 46-55 years. This finding is relatively consistent with Laakso (1997) who finds that the bottom price is found with dwelling aged between 50-70 years. One must remember that Laakso's (1997) study was done over 20 years ago and the development in construction materials and quality may have changed over the years. Our findings are also supported by the fact that dwellings in the 46-55-year category are often undergoing pipe repairs, that are not explicitly included in the model.

Floor number gets a coefficient of 0.015(***) meaning that going up one floor increases the value of the dwelling by around 1.5 % in apartment buildings. Again, comparing to the average priced dwelling in our sample, this turns into an increase of 34.6 euros per square meter. With the average sized dwelling (53.8 sqm), the total price increase is 1860 euros from going one floor up. *Top floor* and *Bottom floor* coefficient estimates get interesting values; -0.007(***) and 0.003(*), respectively. We expected to see a negative coefficient for *Bottom floor* and a positive one for *Top floor* but cannot explain these results. In row houses the *Total floors* coefficient receives an estimate of -0.024(***), indicating that buyers value easy access around the dwelling (everything in one floor). Also, the staircases take away quite a bit of floor area.

Condition affects the price per square meter as expected, but the marginal benefit from upgrading the condition seems to be declining; the relative change between the coefficients decreases from bad to excellent condition. It is also interesting to see that new condition seems to add less value than excellent in row houses. The effect of condition on the dwelling value is considerable; the price difference between bad and excellent condition in a dwelling in apartment building is 29.6 % (coefficient 0.259) on average and 37.0 % in row houses on average (coefficient 0.315***).

Elevator has a 2.4 % increasing effect on the price per square meter on average. A sauna increases the price per square meter by 7.0 % in apartment buildings and 0.9 % in row houses. The difference in the effect of having a sauna between row houses and apartment buildings is fascinating. Over 80

% of row house dwellings have a sauna, but only 30 % of apartment building dwellings have one. Hence, we interpret the result so that a sauna in an apartment building is somewhat a luxury feature that has more demand than supply, but in row houses there are as much saunas as there is demand for them. There are also people that prefer to have no sauna and are willing to pay for not having one.

Being located near to a shore seems to add value to the dwelling. Closeness to a shore increases the value of a dwelling by 9.6 % (apartment buildings) and 19.7 % (row houses) on average, compared to dwellings that are away from shore (ceteris paribus).

Free-hold plot has a price impact of 6.0 % (apartment buildings) and 4.5 % (row houses) on price per square meter. The results are consistent with Tyvimaa, Gibler & Zahirovic-Herbert (2015) who find around 5 % premium on free-hold plot dwellings in HMA. Janssen (2003) finds around 10 - 14 % premium in Stockholm.

The fact that the dwelling is a new built seems to have also effect on the price per square meter, even when we control for the condition and age. New built dwelling has a 4.5 % (apartment buildings) and 5.2 % (row houses) positive impact on price per square meter. The buyer of a new built dwelling often has the opportunity to affect the interior design and colors among other things and is willing to pay a premium on such freedom of choice.

The rented dwelling coefficient gets very small negative, although statistically significant values. One would expect to see negative effect if the dwelling is rented as the new resident has to wait for the previous dweller to move out. Also, owner occupiers may take better care for the dwelling than people living on rented dwellings.

6.4 Linearity of the effect of maintenance charges

We also answer to our research question 4 by studying if the effect of maintenance charge is stable regardless of the level of the charge. In other words, we study, if a one euro change in the maintenance charge affects the price of a dwelling by equal amount regardless of whether the change is from 7 to 8 euros or from 2 to 3 euros, for example. We do this by implementing interaction terms between the continuous *Maintenance charge per square meter* and *Maintenance charge per square meter* dummy variables. Model (5) uses the same variables as model (3) and model (6) uses the same variables as model (4), but additional interaction terms presented in Table 5 are added to the models. We find no statistical significance in the interaction terms, which implies that the effect is constant, and thus, it is equally beneficial to save one euro per square meter in the maintenance charge regardless of the level of the charge. The regression results are shown in Table 5. Control variables are omitted form the table.

Table 5 – Maintenance charge per square meter dummy – Maintenance charge per square meter (continuous) interaction regressions

Table 5 shows the *Maintenance charge per square meter* dummy – *Maintenance charge per square meter* (continuous) interaction regression results. Model (5) is based on the model (3) and model (6) on the model (4), but additional interaction terms are added to those models. Control variables are omitted from the table. All models have postal code and year level fixed effect dummies included. Also, standard errors are clustered at postal code and year level. Standard errors are show in parentheses below the coefficient estimates. ***, ** and * note statistical significance at 1 %, 5 % and 10 % levels, respectively.

	Dependent variable:	
	Ln(Price per sqm)	
	(5)	(6)
Data	Apartment building	Row house
Maintenance charge per sqm	-0.002	-0.020***
	(0.023)	(0.006)
Maint. charge dummy 1 to 2 euros - Maint. charge per sqm (continuous) interaction	-0.022	0.004
	(0.021)	(0.004)
Maint. charge dummy 2 to 3 euros - Maint. charge per sqm (continuous) interaction	-0.027	0.004
	(0.021)	(0.005)
Maint. charge dummy 3 to 4 euros - Maint. charge per sqm (continuous interaction	-0.029	0.004
	(0.021)	(0.005)
Maint. charge dummy 4 to 5 euros - Maint. charge per sqm (continuous) interaction	-0.021	0.008
	(0.022)	(0.005)
Maint. charge dummy 5 to 6 euros - Maint. charge per sqm (continuous) interaction	-0.014	0.007
	(0.022)	(0.005)
Maint. charge dummy 6 to 7 euros - Maint. charge per sqm (continuous) interaction	-0.014	0.008
	(0.022)	(0.007)
Maint. charge dummy 7 to 8 euros - Maint. charge per sqm (continuous) interaction	-0.013	0.011*
	(0.022)	(0.005)
Maint. charge dummy 8 to 9 euros - Maint. charge per sqm (continuous) interaction	-0.016	0.007
	(0.022)	(0.007)
Maint. charge dummy 9 to 10 euros - Maint. charge per sqm (continuous) interaction	-0.016	0.010
	(0.023)	(0.012)
Maint. charge dummy 10 to 20 euros - Maint. charge per sqm (continuous) interaction	-0.015	0.015
	(0.022)	(0.010)
Year dummies	Yes	Yes

Postal code dummies	Yes	Yes
Observations	501,217	172,145
R^2	0.907	0.888
Adjusted R^2	0.907	0.887
Residual Std. Error	0.167 (df = 500100)	0.127 (df = 170736)
Note:		*p**p***p<0.01

6.5 Locational differences in the effect of maintenance charges

In this section we answer our research question 3. We study whether the effect of maintenance charges on dwelling prices is equal across Finland, or if there are some locational differences. We do this again by implementing interaction regression models. The interaction terms are established between locational dummy variables and the *Maintenance charge per square meter* variable. Models (7) and (8) use apartment building data and models (9) and (10) row house data. Models (7) and (8) use the same variables as model (3) and models (9) and (10) use the same variables as model (4), but every model adds the interaction terms presented in Table 6 to the models. We use municipality level interaction terms, the largest municipality being on the top and smallest on the bottom (models (7) and (9)). We also form the locational dummy variables as a group of the largest municipalities in Finland (models (8) and (10)). HMA includes Helsinki, Espoo, Kauniainen and Vantaa, whereas the "Other large municipalities" includes all other municipalities having a population over 100 000 (Tampere, Oulu, Turku, Jyväskylä, Lahti and Kuopio). Control variables are omitted form the table.

Table 6 – Location interaction regressions

Table 6 shows the locational interaction regression results of models (7), (8), (9) and (10). Models (7) and (8) use apartment building data, and models (9) and (10) row house data. Interaction terms are formulated between municipality dummies and *Maintenance charge per square meter*. Espoo and Kauniainen are combined into one entity, HMA consists of Helsinki, Espoo, Kauniainen and Vantaa, and finally, "Other large municipalities" consists of Tampere, Oulu, Turku, Jyväskylä, Lahti and Kuopio. We use the same models as in our main regressions (model (3) for apartment buildings and (4) for row houses) for other parts besides the interaction terms that are added to the models (7), (8), (9) and (10). Other control variables are omitted form the table. Also, standard errors are clustered at postal code and year level. Standard errors are show in parentheses below the coefficient estimates. ***, ** and * note statistical significance at 1 %, 5 % and 10 % levels, respectively.

	Dependent variable:			
	Ln(Price per sqm)			
	(7)	(8)	(9)	(10)
Data	Apartm. building	Apartm. building	Row house	Row house
Maintenance charge per sqm	-0.066***	-0.065***	-0.020***	-0.020***
	(0.008)	(0.008)	(0.002)	(0.002)
Helsinki - Maintenance charge per sqm interaction	0.084***		0.027***	
	(0.011)		(0.005)	
Espoo & Kauniainen - Maintenance charge per sqm interaction	0.076***		0.031***	
	(0.011)		(0.007)	
Vantaa - Maintenance charge per sqm interaction	0.058***		0.025***	
	(0.008)		(0.005)	
Tampere - Maintenance charge per sqm interaction	0.057***		0.028***	
	(0.009)		(0.006)	
Oulu - Maintenance charge per sqm interaction	0.025		-0.036***	
	(0.018)		(0.011)	
Turku - Maintenance charge per sqm interaction	0.051***		0.016**	
	(0.012)		(0.007)	
Jyväskylä - Maintenance charge per sqm interaction	0.020		0.012	
	(0.012)		(0.009)	
Lahti - Maintenance charge per sqm interaction	0.0005		-0.005	
	(0.010)		(0.007)	
Kuopio - Maintenance charge per sqm interaction	0.018		0.002	
	(0.016)		(0.006)	
HMA - Maintenance charge per sqm interaction		0.080***		0.028***
		(0.010)		(0.005)
Other large municipalities - Maintenance charge per sqm interaction		0.040***		0.006
		(0.009)		(0.005)
Year dummies	Yes	Yes	Yes	Yes
Postal code dummies	Yes	Yes	Yes	Yes
Observations	501,217	501,217	172,145	172,145

R^2	0.910	0.910	0.888	0.887
Adjusted R^2	0.910	0.910	0.887	0.886
Residual Std. Error	0.164 (df = 500100)	0.164 (df = 500107)	0.127 (df = 170736)	0.127 (df = 170743)
Note:				*p**p***p<0.01

We find some interesting results. Generally speaking, in the largest municipalities the maintenance charges seem to have very little or no effect on the dwelling prices. In the smaller municipalities the effect is significantly larger and the implied discount rates for the maintenance charges seem to be in line with previous literature. First, we discuss the models (8) and (10) that use HMA and "Other large municipalities" interaction terms and then we proceed to examine models (7) and (9) that use individual municipality level interaction terms. The interpretation of the interaction terms is as follows: the sum of the interaction coefficient and the basic *Maintenance charge per square meter* coefficient shows the total effect in the municipality meaning that the higher (positive) the interaction coefficient is, the less effect the maintenance charges actually have on dwelling prices in that specific municipality.

Models (8) (apartment buildings) and (10) (row houses) combine the separate municipalities into two groups, HMA (Helsinki, Espoo, Vantaa and Kauniainen) and "Other large municipalities". The results from these models follow the results from models (7) and (9) and imply that in the HMA the effect of maintenance charges on dwelling prices is opposite to what would rationally be expected; According to the results, people would be actually willing to pay more for dwellings having higher maintenance charges per square meter than for the low maintenance charge per square meter dwellings. In HMA a one euro increase in the maintenance charge per square meter increases the price per square meter around 1.5 % in apartment buildings and 0.8 % in row houses (interaction coefficient + maintenance charge coefficient). A positive effect means that in practice, there is no discounting of future maintenance charges. In "Other large municipalities" the total effect stays negative (-2.5 % for apartment buildings and -1.4 % for row houses (interaction coefficient + top line maintenance charge coefficient)) meaning that the maintenance charges affect the dwelling prices negatively. In the "Other large municipalities" the implied discount rates are 21 % and 46 % in apartment buildings and row houses, respectively.

Next, we discuss the largest municipalities individually. The interaction coefficients are largest (positive) in Helsinki and Espoo (including Kauniainen) in both apartment buildings (model (7)) and row houses (model (9)). In Helsinki and Espoo, the total effect of maintenance charges on dwelling prices (maintenance charge coefficient + interaction coefficient) is again positive meaning no capitalization of the charges. Also, Vantaa, Tampere and Turku show notably high (positive) interaction coefficients in both apartment buildings (model (7)) and row houses (model (9)) that imply very little capitalization. The results from all of these largest municipalities conflict with the capitalization theory and imply even negative discount rates. In practice, the buyers do not seem to discount the future maintenance charges to be reflected on the dwelling price in these areas. Interestingly, Oulu's interaction coefficients are lower than the ones of Turku even though Oulu's the population is greater. Jyväskylä, Lahti and Kuopio, do not get significant coefficients estimates.

Lastly, we discuss the observations in small municipalities having less than 100 000 residents (that are represented in the top row coefficients). Models (7) and (8) show that the effect is particularly strong in apartment buildings in small municipalities as the price effect from a one euro increase in the maintenance charge is around -6.3 % (coefficient around -0.065***). In row houses (models (9) and (10)), the corresponding price effect is around -2.0 % (coefficients -0.020***) in small municipalities. If we compare the price effect in small municipalities to the regressions results of models (3) and (4) shown in Table 4, we find large differences. In regressions (3) and (4) the effect was shown to be -1.2 % on average in Finland in both, apartment buildings and row houses. The difference between the results from models (3) and (4) and the results observed here in the small municipalities (top line coefficients) shows the magnitude of the impact that the large municipalities have on our regression results. In smaller municipalities the maintenance charges are priced to great extent (at least in apartment buildings it seems) according to the capitalization theory, but the effect is offset by a few large municipalities. The effect observed with apartment buildings in small municipalities is in the same ballpark compared to what is found in the literature of the capitalization of operational costs on dwelling prices. The implied discount rates used in the capitalization of the maintenance charges to dwelling prices for apartment buildings and row houses in small municipalities are around 8.3 % and 33 % according to the results (infinite discount period). The result observed in apartment buildings in small municipalities is in line with the 2 - 10 % implied discount rates observed by Longstreth et al. (1985), Dinan & Miranowski (1986), Janssen (2003) and Kahn & Kok (2014) in their studies of operational cost capitalizations, for example.

The largest municipalities, where the negative effects of maintenance charges on dwelling prices are least prominent, are also often areas where migration is positive and thus demand for housing chronically exceeds the supply. Because these areas are so desired, buyers might not have as much time to consider and evaluate the dwellings as elsewhere; the purchase must be made with incomplete information to secure the dwelling. In these areas the maintenance charges may be overlooked compared to other relevant characteristics (for more rationalization see section *7 – Discussion*). We test if the locational market's ”hotness” could explain our results next.

6.5.1 Market's ”hotness” and the effect on maintenance charges

In hot market areas buyers may have pressure to make quick purchases of dwellings as the demand for dwellings exceeds the supply. When the demand exceeds the supply, housing prices will subsequently rise. Thus, we test whether the effect of maintenance charges on dwelling prices is generally lower in the municipalities of rising dwelling prices. Because in some small municipalities the number of dwelling transactions executed is particularly low, the annual volatility in prices is high, and thus, the average prices per square meter might not be robust. That is why we use the geometric mean of rolling three-year price development per municipality as a proxy for the market hotness in all municipalities. If the geometric mean is positive (i.e. prices have risen), the “Hot market” -dummy gets value 1, and if zero or negative, it gets value 0. The interaction term is formulated between this *Hot market dummy* and the *Maintenance charge per square meter* variable. The results are not comparable to other regressions presented in this thesis as Statistics Finland only provides these price development statistics from 2006 onwards, and there are also other shortcomings in the data in small municipalities. Thus, the sample size is smaller than in other regressions. However, we can still examine the nature of the phenomenon with this data. The results are shown below in Table 7. The models (11) and (12) use the same variables as models (3) and (4) but add the interaction terms and hot market dummies presented in Table 7 to the models. Control variables are omitted form the table.

Table 7 – Hot market interaction regressions

This table shows the hot market interaction regression results. Model (11) uses apartment building data and model (12) uses row house data. Interaction term is formulated between *Hot market dummy* variable and *Maintenance charge per square meter*. Control variables are omitted form the table. Standard errors are clustered at postal code and year level. Standard errors are show in parentheses below the coefficient estimates. ***, ** and * note statistical significance at 1 %, 5 % and 10 % levels, respectively.

	Dependent variable:	
	Ln(Price per sqm)	
	(11)	(12)
Data	Apartment building	Row house
Maintenance charge per sqm	-0.054***	-0.026***
	(0.008)	(0.003)
Hot market - Maintenance charge per sqm interaction	0.042***	0.020***
	(0.008)	(0.003)
Hot market dummy	-0.057**	-0.019**
	(0.024)	(0.007)
Year dummies	Yes	Yes
Postal code dummies	Yes	Yes
Observations	328,960	113,626
R^2	0.908	0.883
Adjusted R^2	0.907	0.882
Residual Std. Error	0.167 (df = 327924)	0.126 (df = 112332)
Note:		*p**p***p<0.01

As the interaction coefficients are positive and statistically significant at 1 % level, we can say that generally speaking, the effect that the maintenance charges have on dwelling prices is indeed smaller in locations of rising dwelling prices. As the interaction coefficient is higher in apartment buildings (model (11)) than in row houses (model (12)), the additional impact that a hot market location has on the effect of maintenance charges on dwelling prices seems to be stronger in apartment buildings than in row houses. The total effects of one euro increases in the maintenance charges on dwelling prices in hot market areas are -1.2 % and -0.6 % in apartment buildings and row houses, respectively. The implied discount rates are 39 % and 100 % calculated from the geometric average selling prices per square meter (2581.9 euros for apartment buildings and 1999.7 euros for row houses in this sub dataset), when discounting to perpetuity. The corresponding implied discount rates for cold market areas are 8.8 % and 23 %. These discount rates are in the same ballpark as discount rates calculated above in section *6.5 – Locational differences in the effect on maintenance charges*; in hot market areas or large municipalities buyers do not take in account the maintenance charges as much as in the cold market areas or small municipalities. The hot market areas are typically also the largest municipalities and cold areas typically the smallest, so we cannot conclude whether the observations actually are a result from the market hotness or from some other characteristic that is associated with larger municipalities and omitted from the model. To better asses the market hotness's/dwelling hotness's impact on the capitalization of maintenance charges, we construct sales time dummy interaction terms based on individual dwellings' sales times rather than on a municipality level.

6.5.2 Sales time and the effect of maintenance charges

The general dwelling price development is a good indicator of the areal hotness of the market, but at the same time the areas where prices are on the rise are the biggest municipalities making it hard to know if the effects are a result from the price development or from other characteristics that are common in these municipalities. By doing the market hotness regression with actual sales times for every single dwelling we are taking a different angle to this matter as the sales times are not tied to the municipality borders.

Some dwellings in particularly hot locations where the supply is limited are sold immediately as they come to the market. Some dwellings, in turn, are marketed for months, or even years. This is often the case in locations far from major growth centers such as HMA, Turku, Tampere or Oulu as the Figure 3 in section *2.1.1 – Characteristics of the Finnish housing market* showed us. However, there are dwellings that are not sold like hot cakes in major cities, and vice versa, some dwellings are sold quickly regardless of the outlying location. We expect that the quicker the individual dwelling is sold, the less the buyer pays attention to the level of the maintenance charge as the buyer probably has had a hurry to buy the dwelling so that someone else does not get it. Our regression results support our expectations.

We employ *Sales time* dummy variables for every 25 days in sales time for each dwelling transaction. For example, if the dwelling is sold in 116 days, the *Sales time 100 to 125 days* dummy gets value 1 for that specific dwelling, and other dummies 0. These dummies are inserted in the regression models (13) (for apartment building data) and (14) (for row house data) individually as additional control variables and in interaction terms with the *Maintenance charge per square meter* variable. For other parts, the models (13) and (14) follow the main regression models (3) and (4), respectively. The interaction coefficients of model (13) are plotted in Figure 14 (apartment building data), and of model

(14) in Figure 15 (row house data). Sales time between 0 - 25 days is the reference level in the regressions. Both figures show a steadily sloping line; the longer the sales time, the smaller the interaction coefficient.

We did not expect to see any positive coefficients, because the reference level is the shortest sales time, but for some reason the interaction coefficients of *Sales time 25 to 50 days* are still positive. However, the overall results are clear as the effect gets inevitably stronger with the increasing sales times. We conclude that if an individual dwelling is "hot", (defined by short sales time) the maintenance charge is not taken in account and does not reflect to the dwelling price as strongly compared to when the dwelling is "cold". The result is consistent with our findings in section *6.5.1 – Market's "hotness" and the effect on maintenance charges*.

Figure 14 – Sales time dummy – Maintenance charge per square meter interaction coefficients – apartment building data

Figure 14 presents the *Sales time* dummy interaction regression results from the model (13). The coefficient values represent the additional impact that a particular sales time has on the capitalization of maintenance charges to dwelling prices. The regression model (13) follows the model (3) but has also additional *Sales time* dummy control variables as well as the interaction terms included. Standard errors are clustered at postal code and year level. The candles represent the 95 % confidence intervals.

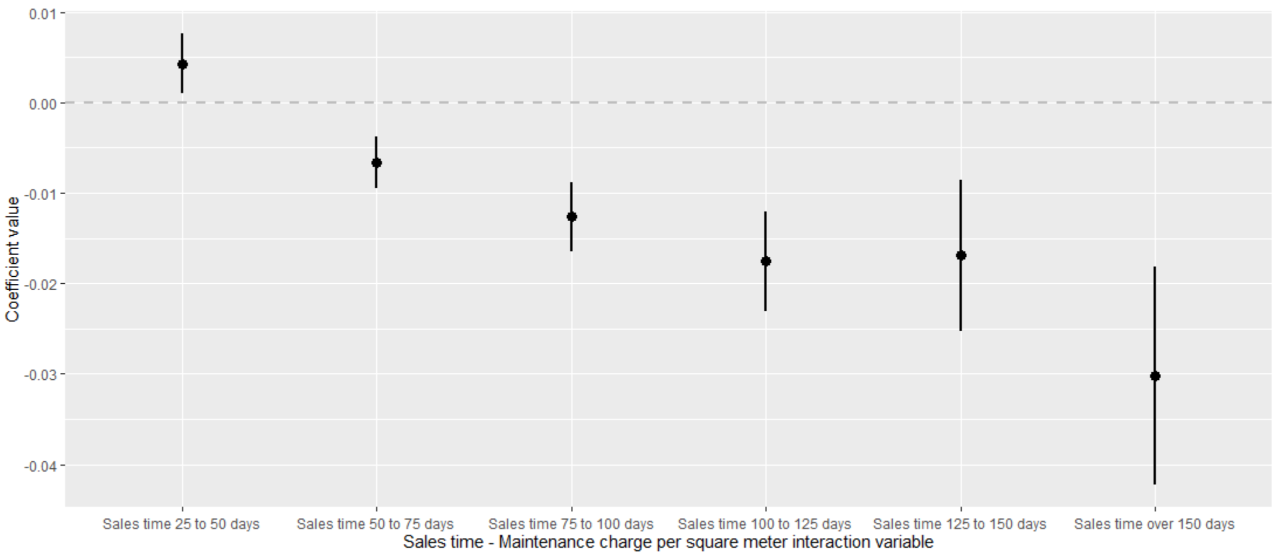
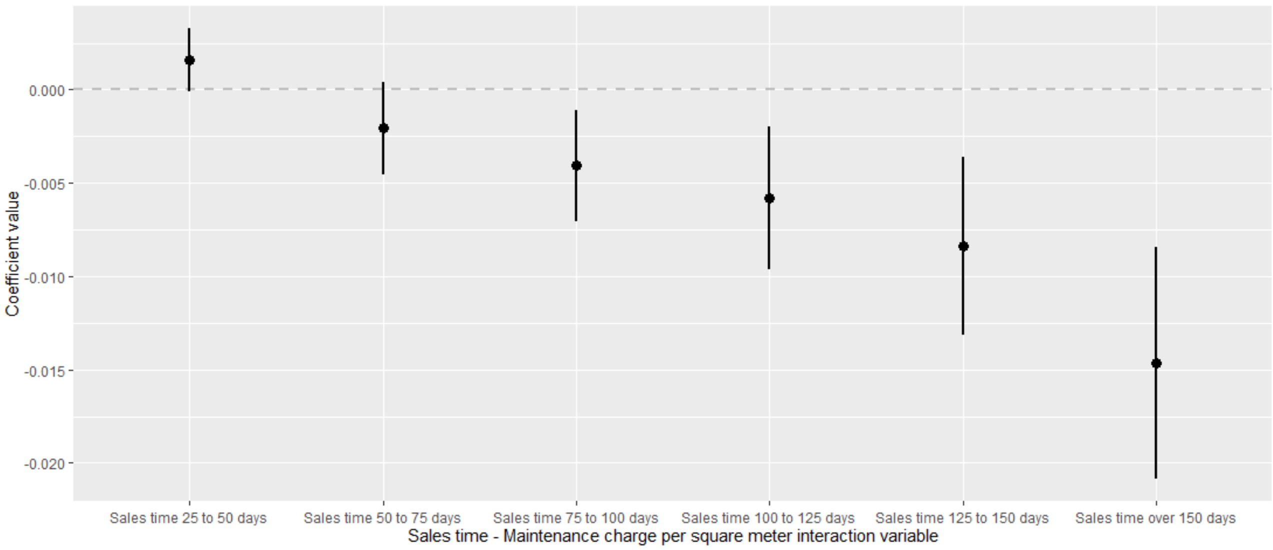


Figure 15 – Sales time dummy – Maintenance charge per square meter interaction coefficients – row house data

Figure 15 presents the Sales time dummy interaction regression results from the model (14). The coefficient values represent the additional impact that a particular sales time has on the capitalization of maintenance charges to dwelling prices. The regression model (14) follows the model (4) but has also additional Sales time dummy control variables as well as the interaction terms included. Standard errors are clustered at postal code and year level. The candles represent the 95 % confidence intervals.



6.6 Area – Maintenance charge per square meter interaction regression

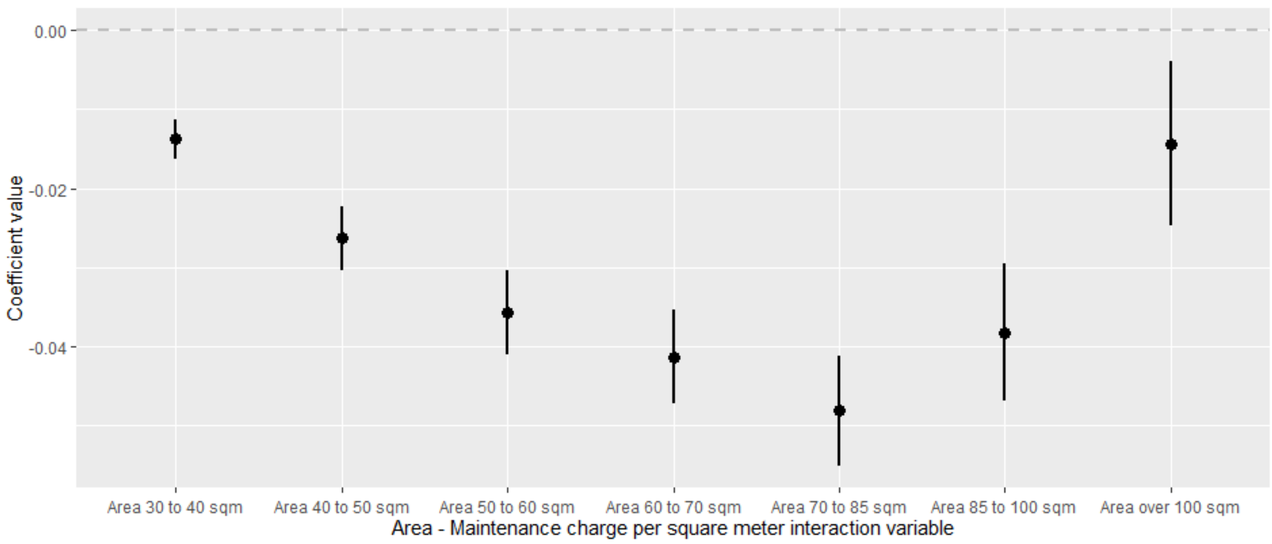
Next, we take on our research question 5. We believe that the maintenance charges might be capitalized differently depending on the absolute size of the dwellings. The rationale is that in large dwellings a change in the per square meter maintenance charge affects the total charge significantly. Thus, people might take the changes in account more carefully as they have more impact on their financials. In literature, quite small personal discount rates are found with small sums of money as we discussed in sections *1.1 – Background and motivation* and *3.2 – Studies on housing prices and operating expenses*.

We construct *Area – Maintenance charge per square meter* interaction terms and include them in the original model (3) forming a new model (15) to examine the additional effect that different areas bring. The *Area* dummies used in the interaction terms are set as follows: 30-40, 40-50, 50-60, 60-70, 70-85, 85-100 and over 100 square meters. The interaction term coefficients now represent the additional effect that a one euro increase in the maintenance charge has on the dwelling price in a certain dwelling area compared to the effect of the maintenance charge at the reference level, i.e.

within under 30 square meter dwellings. Figure 16 presents the interaction coefficients of apartment building data regression. We can see that the trend is downward sloping meaning that the capitalization of the maintenance charges differs among different sized dwellings. If the area of the dwelling had no impact on the capitalization, the interaction coefficients would get a value of zero. Now, as we move towards larger dwellings in which a change in the maintenance charge per square meter affects the total charge significantly, we observe that an increase in the maintenance charge per square meter has a stronger negative effect on the dwelling price.

Figure 16 – Area – Maintenance charge per square meter interaction coefficients

Figure 16 shows the *Area – Maintenance charge per square meter* interaction coefficients from model (15). The model follows the model (3) but has additional *Area – Maintenance charge per square meter* interaction terms also included. This model uses apartment building data and has postal code and year level fixed effect dummies. The reference level is the dwelling size between 0 - 30 square meters. Also, standard errors are clustered at postal code and year level. Candle lines represent 95 percent confidence intervals.



We interpret the result like this: The real-life difference to the buyer's financials from changing the maintenance charge per square meter in a small dwelling is minor (e.g. going from 70 euros to 80 euros) and thus no major impact on the price per square meter is observed. With larger dwellings that have also larger maintenance charges (in total euro terms) the effect from increasing the maintenance charge per square meter is stronger, as the change starts to make a difference (e.g. going from 700 euros to 800 euros), and thus the effect is seen as a larger decrease in the price per square meter. We remind that we are talking about equal sized increases in the per square meter maintenance charges here (e.g. 1 euro increase), but when the area increases, the total maintenance charge also increases, and thus we see such results. However, the converting trend at larger dwelling sizes (seen in far right on Figure 16) does not follow the pattern. We are not able to specify whether this effect comes from model specification, small sample sizes or omitted variables. We test correlations between variables but find none that would be alarming and explain the curve. We also look at the larger dwellings in detail in order to find some patterns that would explain the phenomenon but find none. It is also possible, that the effect is not a result from mis-specified model/small sample, but really exists. We also test the same interaction terms with row house data, but the results are not significant, and we decide not to report them.

6.7 Other interesting results

In addition to the results presented above that answered to our research questions we perform some additional analyses. The most interesting findings relate to capitalization of maintenance charges when controlling for the dwelling price. In other words, we test if the maintenance charges are taken in account similarly in cheap and expensive dwellings. On a general level, we find that the capitalization of the maintenance charges increases as the total price of the dwelling increases. The analysis in conducted with Total price dummy – Maintenance charge per square meter interaction terms that we add to the models (3) and (4) producing the models (16) and (17). The Total price dummies that are also included in the interaction terms are constructed in following way: 0-100k euros, 100-300k euros, 300-500k euros and over 500k euros. The interaction term describes the additional effect that a one euro change in the maintenance charge has on the price per square meter accounting for the total price of the dwelling. Also, Total price dummies are included in the models as additional control variables. The interaction regression results are shown below in Table 8.

Table 8 – Total price interaction regressions

Table 8 shows the *Total price – Maintenance charge per square meter* interaction term coefficients from models (16) and (17). Model (16) uses apartment building data and model (17) uses row house data. Model (16) follows model (3) while model (17) follows model (4), but additional *Total price* dummies and interaction terms are also included. Other control variables are omitted form the table. All models have postal code and year level fixed effect dummies included. The reference level is the total price under 100 000 euros. Standard errors are clustered at postal code and year level. Standard errors are show in parentheses below the coefficient estimates. ***, ** and * note statistical significance at 1 %, 5 % and 10 % levels, respectively.

	Dependent variable:	
	Ln(Price per sqm)	
	(16)	(17)
Data	Apartment building	Row house
Maintenance charge per sqm	-0.006**	-0.012***
	(0.002)	(0.002)
Total price 100k-300k euros dummy - Maintenance charge per sqm interaction	0.008	0.011***
	(0.006)	(0.004)
Total price 300k-500k euros dummy - Maintenance charge per sqm interaction	-0.020***	0.007
	(0.006)	(0.007)
Total price over 500k euros dummy - Maintenance charge per sqm interaction	-0.024***	-0.016***
	(0.005)	(0.004)
Total price 100k-300k euros dummy	-0.014	0.004
	(0.019)	(0.010)
Total price 300k-500k euros dummy	-0.001	-0.131***
	(0.017)	(0.016)
Total price over 500k euros dummy	0.019	-0.065***
	(0.016)	(0.008)
Year dummies	Yes	Yes
Postal code dummies	Yes	Yes
Observations	501,217	172,145
R^2	0.909	0.896
Adjusted R^2	0.909	0.895
Residual Std. Error	0.165 (df = 500100)	0.122 (df = 170736)
Note:		*p**p***p<0.01

Elinder & Persson (2017) find that property tax reformations are priced correctly in the dwelling prices only in the top 1 % highest priced dwellings. They argue that the people buying these dwellings could be more financially literate to better understand the NPV effects of future costs. We find that the capitalization of maintenance charges seems to increase as the dwellings price increases. Hence, the findings of Elinder & Persson (2017) are somewhat similar with our results, but their explanation seems rather shaky. Based on our results we cannot conclude why we observe such behavior.

7 Discussion

There could be several possible explanations for the results we observe. The market is clearly not efficient when pricing the maintenance charges. The implied discount rates calculated for the maintenance charges from the regressions in sections *6.3 – The observed effect of maintenance charges on dwelling prices* (43 % and 55 % for apartment buildings and row houses, respectively) and *6.5 – Locational differences in the effect of maintenance charges* are a result of several possible behavioral biases, but also rational explanations can be found. The results from *6.6 – Area – Maintenance charge per square meter interaction regression* are in line with our expectations to a certain extent, but some inconsistencies are observed in the results. Next, we go through some of the rational explanations for our results with the help of some behavioral biases observed often in the literature.

One of the most prominent biases represented in the literature is anchoring. Tversky and Kahneman (1974) were one of the first to study this bias. People tend to make estimates by setting starting points (or anchors) and then adjust the starting point to get to the final value. The value of the anchor affects the final answer converting the final estimate towards the anchor. (Tversky and Kahneman, 1974.) For example, when pricing maintenance charges to dwelling prices, people could use the prices of similar dwellings in the same neighborhood as reference points even if they have differing maintenance charges which would lead to observing minor capitalization of the maintenance charges on dwelling prices, much like our results show. This might also explain the locational differences. In larger municipalities the transaction volume is high, and it is relatively easy to compare a dwelling price to other similar dwellings' prices nearby, whereas in the small municipalities it might be difficult to find comparable transactions to which anchor the price. For this reason, it might be the case that in larger municipalities, where the anchoring is easier, the capitalization of maintenance charges on dwelling prices seems to be smaller.

The anchoring is accompanied with the focusing effect that was especially studied by Schkade and Kahneman (1998). People can only focus on a limited number of attributes at a time and some attributed will always be disregarded. Buying a dwelling is a complex process. These processes require special knowledge and often outside help to translate aspects of the purchase in terms that the buyers can understand (Thaler and Sunstein, 2009). In such complex processes the focus is often aimed to the most important aspects that are usually related to the visible dwelling characteristics, such as location, area and condition. The maintenance charges are generally displayed in total amounts (e.g. 250 euros per month) and not in per square meter terms making the valuation even harder. Thus, it is easy to disregard the maintenance charges when valuing dwellings and price them incorrectly.

The availability bias relates to the former; people tend to use information that is readily available rather than produce the information themselves. What affects the availability of the information is its vividness. Thinking about the characteristics of dwellings, the most vivid are the visible attributes such as the condition. Characteristics such as electricity consumption or water expenditures and maintenance charges are harder to observe, or at least they are not presented in the same vivid manner as the positive or more important characteristics. The fact that the maintenance charges are presented in total euros per month terms rather than per square meter terms further takes away the focus.

People usually manage poorly with decisions that require the postponement of benefits and involve immediate costs (Sustain, 2014). Basically, this implies that consumers tend to have high personal discount rates, which has also been found by other researchers such as Thaler (1981). If you manage to find a dwelling that fits your needs almost perfectly, it is tempting to disregard the negative aspects, such as future maintenance charges. The maintenance charges are realized over future years, but the benefits of the perfect dwelling are realized immediately. The discount period for maintenance charges is very long, even if you only discount the charges for the expected useful life span of the dwelling (e.g. 50 years for new built dwellings) making it very easy to completely disregard the costs occurring far in the future. We conclude that all these biases discussed above are probably present in the Finnish housing market and have an effect on the pricing of maintenance charges on dwelling prices.

8 Conclusions

Housing plays an important role in the life of Finnish individuals. It accounts for a major part of the average household wealth. Because housing relates to so many aspects, pricing it correctly makes the lives of Finnish individuals just that much easier. Pricing housing correctly makes the housing market more efficient as it increases the market liquidity and consequently sales times are reduced. If an asset is priced incorrectly and either the buyer or the seller does not know the true value of the asset, the spread between the bid and ask prices will increase, eventually to the point that the trading will halt completely. No one wants to make a bad deal. Another example of the benefits of accurate housing pricing could concern applying for a new loan for an investment property, for example: it would be in the best interest of the bank, but also the loan seeker, to have the collateral dwelling valued correctly. Our motivation for this thesis comes from the fact that housing characteristics and their pricing has been studied a lot, but no previous research specifying on the pricing of maintenance charges in Finland has been made, even if it is such an important aspect.

Our research questions are: what is the effect of maintenance charges on dwelling prices and what are the implied discount rates? Does the price effect differ across municipalities? Is the effect of maintenance charges on dwelling prices linear w.r.t the level of the charge? And finally, is the effect stronger in large apartments with larger absolute charges?

8.1 Concluding remarks of the results

The key finding in this thesis is that dwellings are overpriced w.r.t maintenance charges in Finland on average. The implied discount rates observed are very high compared to rates found in papers studying operational cost capitalization to dwelling prices. The rates are also very high when comparing to the required rates of return of investments considered traditionally being rather risky, like stock investments. We do not explicitly denote what would be a "correct" discount rate for the maintenance charges as it cannot be calculated/estimated precisely, but instead leave the assessment of the correctness of the implied discount rates to the reader.

Our main findings in section *6.3 – The observed effect of maintenance charges* indicate that the maintenance charges are not discounted fully to the dwelling prices in neither, apartment buildings nor in row houses in Finland. Our further analysis in section *6.5 – Locational differences in the effect of maintenance charges* indicates that the effect on the level whole Finland stems from the larger municipalities of that represent notable weight in the dataset. In the largest municipalities and especially in their best locations where population growth drives demand for housing above the supply, is the availability of dwellings and their characteristics limited. A purchase needs to be made fast when a suitable dwelling comes to the market; especially in Helsinki the fastest buyer often secures the deal. The high demand for housing in these areas also drives the prices of dwellings up as we have discussed. When people need to make fast decisions that require the consideration of multiple factors, some factors will always be overlooked. It seems that little emphasis is given to the level of maintenance charges in these situations. When we move outside of the HMA and the largest municipalities in Finland, the magnitude of overpricing of dwellings w.r.t maintenance charges reduces rapidly. Furthermore, the locational differences in the capitalization of maintenance charges appear to be much stronger with apartment buildings than with row houses. To translate these results to a single statement we would put it like this: it is in the buyer's best financial interest to buy a dwelling with a small maintenance charge. The value of this advice is highlighted especially among apartment buildings in the largest municipalities.

Some other interesting findings are also made. We test how the hotness of the market affects the pricing of maintenance charges and find that in hot market areas the dwellings are more likely overpriced than in the cold markets. We also test the hotness effect on the individual dwelling level and find similar results. It happens to be that the hot market municipalities are also the biggest municipalities in Finland. Also, the hot dwellings are typically found in the same areas. Thus, we conclude that we cannot state causal effects, but merely report what we observe. Still, this finding supports the advice that when considering dwellings in areas where the demand for housing is high compared to the supply, one should go for the dwellings with low maintenance charges.

We also test how the changes in the maintenance charges affect dwelling pricing depending on the level of maintenance charges in section *6.4 – Linearity of the effect of maintenance charges*, finding that the level of maintenance charge per square meter has no impact on how changes in the maintenance charges are capitalized to dwelling prices (e.g. there is no matter if the charge is 2 euros or 9 euros per square meter, a change in the charge is capitalized similarly). In section *6.6 – Area – Maintenance charge per square meter interaction regression,* we test the following: how the maintenance charges are capitalized to the dwelling prices depending on the level of the total charge (i.e. when taking in account the area of the dwelling). We expect that when the area of the dwelling is large and consequently a change in the maintenance charge per square meter has a large effect on the total maintenance charge expenses (in total euro terms), even a small change in the maintenance charge per square meter would have a large impact on the dwelling valuation compared to dwellings with small areas. In other words, people would take the maintenance charges better into account when valuing a large dwelling with large monthly charges. On some level, we find support for this argument in the apartment buildings even though there seems to be large deviations in the results especially with the largest dwellings.

Lastly, we test how the total dwelling price affects the capitalization of the maintenance charges in section 6.7 – *Other interesting results* finding that in more expensive dwellings the capitalization is stronger. This result might be explained by the financial literacy of people buying expensive dwellings following Elinder & Persson's (2017) argumentation.

Our results are probably a consequence of several behavioral biases as we discussed in section *7 – Discussion*. Our results are somewhat in line with previous findings of typical personal discount rates that tend to be relatively high, but considerably different from studies on capitalization of operational expenses as we find much higher implied discount rates. The behavioral biases are possibly further intensified by the fact that dwelling transactions require lots of expertise and the consideration of multiple factors simultaneously. In these kinds of situations, we often make bad decisions in a purely financial perspective. People buy dwellings usually as a home for themselves, so the heart often triumphs the brain, thus we are not surprised to see such results.

8.2 Evaluation of the study

No research paper is completely flawless, and neither is ours. However, we have tried to eliminate the problems as well as we could. The evaluation of the study focuses on two aspects, the data and the methodology. The main data is provided by the KVKL, to which only professionals get to list their transactions to. Nevertheless, there are lots of mistakes in the raw data. We paid particular attention to the data clean-up process and checked pedantically all the variables in order to eliminate the possible mistakes, but for sure some are still left in the data used. However, the sheer size of the dataset overpowers the effect that some minor mistakes in the data could have on our results. The extensivity of the data also allowed us to study the phenomenon in different areas and dwelling types.

The methodology we have used follows the common standards in the fields of finance and real estate economics. We have used the hedonic regression method and tested several alternative models and analyzed them with respect to the OLS assumptions before choosing the final form of the variables and the model. We have done such an extensive analysis of the model and variable specifications in the methodology section, that we find no reason to state separate robustness checks in this thesis.

The data availability sets limits to our study, and we could not include all the desired variables in the model, even if they would have been rational. Thus, we suffer from some omitted variable biases. The problem is more severe in particularly expensive and cheap dwellings, and for that reason we cut the price per square meter at the lower and upper ends, and leave those extreme cases out form the scope of the study. Regressions regarding row houses suffer from the omitted variable bias more than the ones regarding apartment buildings (e.g. omitted variables considering yards). Probably the most meaningful omitted variable in our regression models is the renovation backlog. Some housing studies also have travel time variables (e.g. time to centrum, to sea, to park), but unfortunately, we were not able to find a free API for that purpose. However, the R-squared values of our regressions lie around the 90 % level which implies that the models explain the price relatively well. The fitted values of our models are systematically close to the observed values as well.

All of our regressions except for the *Area – Maintenance charge per square meter* interaction analysis have consistent results that are in line with our expectations. However, the implied discount rates we find are larger than the ones found in relevant literature related to the components of the maintenance charge. Thus, there is reason for further research. Our analysis also focused, inter alia, on the locational differences, and found that the maintenance charges are capitalized poorly to dwelling prices in larger municipalities. We continued by studying if the phenomenon was linked to the rising dwelling prices and shorter sales times (to which both we refer as market/dwelling hotness) and found that in hot markets the capitalization is weaker and dwellings more overpriced. The problem here is that we cannot say that the reason for observing a weaker capitalization would be the hotness of the market, i.e., our analysis does not prove the causality. It is however worth underlining that our main goal was to study whether the maintenance chargers are fully taken in account in dwelling prices, and eventually we find some mispricings. By finding these mispricings we can enhance the efficiency of the market possibly making it more liquid. Thus, the causality problem in the further analyses does not undermine the study.

8.3 Future research

As the topic of maintenance charges has been studied only cursorily in the literature, there is room for lots of future research. Our research could be elaborated by using a more extensive dataset. If one was able to combine the rich but imprecise dwelling advertisement data (e.g. Oikotie.fi or Etuovi.com) with the less rich but more precise data provided by KVKL, the regression results would be more accurate. Dwelling advertisements contain the housing company level renovation data, but the prices are ask-prices, not realized transaction prices like the ones we have used. One could also study the topic using some Austrian, Dutch or Norwegian data, as they have similar maintenance charge systems.

Like mentioned, our study cannot deduce the causality between the market hotness and the effect of maintenance charges in different municipalities. The future research could look at this phenomenon in more detail to prove the causality, or to find alternative explanations for the locational differences we find. Also, the *Area – Maintenance charge per square meter* interaction regression we run cannot be described robust, and thus that could be examined further.

The last point we have relates to the discount rates. To put it simply we cannot say what would be a rational implied discount rate for the maintenance charges. This is mainly because we do not know how stable the level of maintenance charges is in the end. Even though we can quite safely argue that the implied discount rates for the maintenance charges we find, 43 % for apartment buildings and 55 % for row houses (discounting to perpetuity) in the whole combined dataset for Finland, are far from rational levels, it would be beneficial to have an understanding what the discount rates should be taking in account the volatility of the charges and other matters affecting the discount rates.

9 References

Benzion, U., Rapoport, A., & Yagil, J. (1989). Discount rates inferred from decisions: An experimental study. Management science, 35(3), 270-284.

Can, A. (1992). Specification and estimation of hedonic housing price models. Regional science and urban economics, 22(3), 453-474.

Carson, R. T., & Hanemann, W. M. (2005). Contingent valuation. Handbook of environmental economics, 2, 821-936.

Coase, R. H. (1937) The nature of the firm. Economica, New series, Vol. 4, No. 16 (Nov., 1937), pp. 386-405

Colwell, P. F. (2002). Tweaking the DiPasquale-wheaton model. Journal of Housing Economics, 11(1), 24-39.

Cropper, M. L., Deck, L. B., & McConnell, K. E. (1988). On the choice of funtional form for hedonic price functions. The review of economics and statistics, 668-675.

Dhrymes, P. (1971) Price and Quality Changes in Consumer Capital Goods. In Grilliches, Z. (ed.): Price Indexes and Quality Change. Harvard University Press.

Dinan, T. M., & Miranowski, J. A. (1989). Estimating the implicit price of energy efficiency improvements in the residential housing market: A hedonic approach. Journal of Urban Economics, 25(1), 52-67.

DiPasquale, D., & Wheaton, W. C. (1992). The markets for real estate assets and space: A conceptual framework. Real Estate Economics, 20(2), 181-198.

Einiö, M., Kaustia, M., & Puttonen, V. (2008). Price setting and the reluctance to realize losses in apartment markets. Journal of Economic Psychology, 29(1), 19-34.

Elinder, M., & Persson, L. (2017). House price responses to a national property tax reform. Journal of Economic Behavior & Organization, 144, 18-39.

Farrar, D. E., & Glauber, R. R. (1967). Multicollinearity in regression analysis: the problem revisited. The Review of Economic and Statistics, 92-107.

Fisher, I. (1907). The rate of interest. Garland Pub..

Follain, J. R., & Malpezzi, S. (1980). Dissecting housing value and rent: Estimates of hedonic indexes for thirty-nine large SMSAs (Vol. 249, No. 17). Urban Institute Press.

Goodman, A. C., & Thibodeau, T. G. (1995). Age-related heteroskedasticity in hedonic house price equations. Journal of Housing Research, 25-42.

Goodman, A. C., & Thibodeau, T. G. (1997). Dwelling-age-related heteroskedasticity in hedonic house price equations: An extension. Journal of Housing Research, 299-317.

Gordon, M. J., & Shapiro, E. (1956). Capital equipment analysis: the required rate of profit. Management science, 3(1), 102-110.

Grigsby, W. G. (1963). Housing markets and public policy. University of Pennsylvania Press.

Gu, B., & Hitt, L. (2001). Transaction costs and market efficiency. ICIS 2001 Proceedings, 11.

Halvorsen, R., & Pollakowski, H. O. (1981). Choice of functional form for hedonic price equations. Journal of urban economics, 10(1), 37-49.

Harjunen, O., & Liski, M. (2014). Not so myopic consumers-evidence on capitalization of energy technologies in a housing market.

Janssen, C. T. (2003). Estimating the effect of land leases on prices of inner-city apartment buildings. Urban Studies, 40(10), 2049-2066.

Kahn, M. E., & Kok, N. (2014). The capitalization of green labels in the California housing market. Regional Science and Urban Economics, 47, 25-34.

Knight, J. R., Miceli, T. and Sirmans, C. F., 2000. Repair expenses, selling contracts and house prices. Journal of Real Estate Research, Vol. 20, Issue 3, pp. 323-336.

Kuminoff, N. V., Parmeter, C. F., & Pope, J. C. (2010). Which hedonic models can we trust to recover the marginal willingness to pay for environmental amenities?. Journal of environmental economics and management, 60(3), 145-160.

Laakso, S. (1997). Urban housing prices and the demand for housing characteristics. A study on housing prices and the willingness to pay for housing characteristics and local public goods in the Helsinki Metropolitan Area. ETLA A.

Lancaster, K. J. (1966). A new approach to consumer theory. Journal of political economy, 74(2), 132-157.

Longstreth, M., Coveney, A. R., & Bowers, J. S. (1985). The effects of changes in implicit energy costs on housing prices. Journal of Consumer Affairs, 19(1), 57-73

Mahlon, R. (1975). An Econometric Analysis of the Urban Housing Market.

Malpezzi, S., 2003. Hedonic pricing models: a selective and applied review. Housing Economics: Essays in Honor of Duncan Maclennan, Edited by K. Gibb and A. O'Sullivan. Blackwell.

Molin, E.J.E. (2011) Conjoint Analysis. In Jansen, S.J.T., Coolen, H.C.C.H. and Goetgeluk, R.W. (eds.) The Measurement and Analysis of Housing Preferences and Choices. Springer, Dordrecht.

O'Sullivan, A., 1996. Urban Economics, 3rd ed., Times Mirror Higher Education Group.

Oates, W. E. (1969). The effects of property taxes and local public spending on property values: An empirical study of tax capitalization and the Tiebout hypothesis. Journal of political economy, 77(6), 957-971.

Oikarinen, E. (2014). Studies on housing price dynamics.

Palmquist, R. B. (1984). Estimating the Demand for the Characteristics of Housing. The Review of Economics and Statistics, 394-404.

Parr, J. B. (2007). Spatial definitions of the city: four perspectives. Urban Studies, 44(2), 381-392.

Peltola, R. (2011, May). Leasehold discount in dwelling prices. A neglected view to the challenges facing.

Porter, M. E. (1989). How competitive forces shape strategy. In Readings in strategic management (pp. 133-143). Palgrave, London.

Rosen, S. (1974). Hedonic prices and implicit markets: product differentiation in pure competition. Journal of political economy, 82(1), 34-55.

Rothenberg, J., Galster, G. C., Butler, R. V., & Pitkin, J. R. (1991). The maze of urban housing markets: Theory, evidence, and policy. University of Chicago Press.

Schkade, D. A., & Kahneman, D. (1998). Does living in California make people happy? A focusing illusion in judgments of life satisfaction. Psychological science, 9(5), 340-346.

Sirmans, S., Macpherson, D., & Zietz, E. (2005). The composition of hedonic pricing models. Journal of real estate literature, 13(1), 1-44.

Straszheim, M. R. (1975). An Econometric Analysis of the Urban Housing Market NBER Columbia University Press, New York.

Sunstein, C. R. (2014). Why nudge?. Yale university press.

Thaler, R. H., & Sunstein, C. R. Nudge: Improving decisions about health, wealth, and happiness.

Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. science, 185(4157), 1124-1131.

Tyvimaa, T., Gibler, K. M., & Zahirovic-Herbert, V. (2015). The effect of ground leases on house prices in Helsinki. Journal of Housing and the Built Environment, 30(3), 451-470.

Virén, M. (2011). Does housing allowance feed through into rental prices?. Bank of Finland Research Discussion Paper, (11).

Wilhelmsson, M. (2008). House price depreciation rates and level of maintenance. Journal of Housing Economics, 17(1), 88-101.

Database sources:

Statistics Finland databases *StatFin* and *Municipality key figures*

Federation of Real Estate Agency (Kiinteistönvälitysalan keskusliitto), Hintaseurantapalvelu (HSP)

Other references:

Limited Liability Housing Companies Act (1599/2009; amendments up to 547/2010 included; asunto-osakeyhtiölaki

10 Appendix

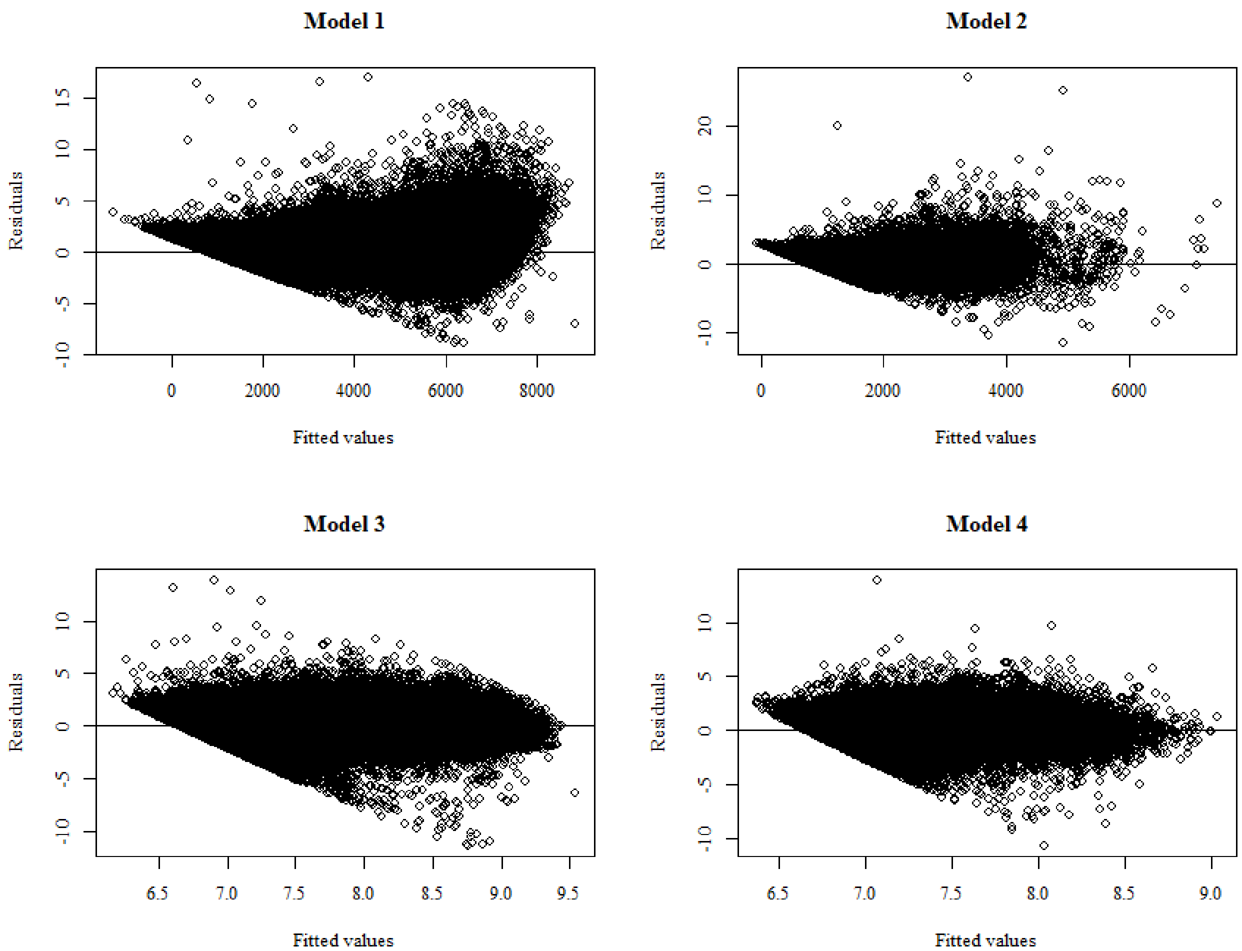
Appendix 1 - Correlation matrix of the complete dataset – selected variables

Appendix 1 shows the pairwise correlations between the variables used in our main regression models. Variables *Age 11 to 15, Age 16 to 25, Age 26 to 35, Age 36 to 45, Age 46 to 55, Age 56 to 65, Age 66 to 80* and *Room number 2*, *Room number 3* and *Room number 4* are hidden form the table for sake of space saving. However, none of the hidden variables show alarming correlations.

	Maint. charge per sqm	Age 0 to 5	Age 6 to 10	...	Age 81 to 95	Age over 95	Ln(Area)	Room number 1	...	Room number 5	Room number 6 or more	Floor number	Total floors	Bottom floor	Top floor	Condition bad	Condition satisfactory	Condition good	Condition excellent	Condition new	Free hold plot	Rented dwelling	Elevator	Sauna	Shore	New built dwelling
Maintenance charge per sqm	1			...																						
Age 0 to 5	-0.05	1		...																						
Age 6 to 10	-0.07	-0.1	1	...																						
...	...	...	...	...																						
Age 81 to 95	0.1	-0.07	-0.03	...	1																					
Age over 95	0.08	-0.05	-0.03	...	-0.02	1																				
Ln(Area)	-0.25	0.01	0.07	...	-0.05	0	1																			
Room number 1	0.15	-0.03	-0.06	...	0.1	0.05	-0.7	1																		
...	...	...	...	...	...	...	...	...	...																	
Room number 5	-0.05	0	0.02	...	0	0.01	0.28	-0.07	...	1																
Room number 6 or more	-0.02	-0.02	-0.01	...	0.01	0.03	0.16	-0.03	...	-0.01	1															
Floor number	0.17	0.07	-0.03	...	0.07	0.02	-0.17	0.09	...	-0.05	-0.01	1														
Total floors	0.23	0.1	-0.03	...	0.09	0.02	-0.23	0.13	...	-0.06	-0.02	0.7	1													
Bottom floor	0.1	-0.05	-0.03	...	0	0.01	-0.13	0.07	...	-0.04	-0.02	-0.36	0	1												
Top floor	0.11	-0.05	-0.04	...	0.01	0.02	-0.09	0.04	...	-0.02	-0.01	0.41	0	-0.19	1											
Condition bad	0.03	-0.07	-0.03	...	0.05	0.03	-0.05	0.05	...	0	0.01	0.03	0.05	0.02	0.01	1										
Condition satisfactory	0.08	-0.26	-0.12	...	0.02	0.01	-0.12	0.11	...	0	0.01	0.05	0.07	0.05	0.03	-0.09	1									
Condition good	-0.15	0.08	0.13	...	-0.03	-0.01	0.17	-0.15	...	0.01	-0.01	-0.09	-0.13	-0.05	-0.03	-0.22	-0.85	1								
Condition excellent	0.06	0.15	0.01	...	0	0	-0.05	0.03	...	-0.01	0	0.04	0.06	0	-0.01	-0.02	-0.06	-0.15	1							
Condition new	0.12	0.39	-0.04	...	-0.02	-0.01	-0.08	0.05	...	-0.01	-0.01	0.08	0.1	-0.02	-0.02	-0.03	-0.11	-0.27	-0.02	1						
Free hold plot	-0.09	-0.08	-0.01	...	0.05	0.05	0.02	0.01	...	0.01	0.02	0.04	0.06	-0.01	-0.01	0.01	0.04	0	-0.02	-0.09	1					
Rented dwelling	0.02	0.01	-0.01	...	0.01	0	-0.13	0.13	...	-0.02	-0.01	0.03	0.05	0.02	0.02	0	0.03	-0.01	-0.01	-0.05	0	1				
Elevator	0.12	0.21	0.02	...	0.05	-0.02	-0.16	0.08	...	-0.05	-0.02	0.48	0.66	-0.02	-0.02	0.03	0	-0.05	0.09	0.08	0.06	0.03	1			
Sauna	-0.28	0.28	0.19	...	-0.11	-0.07	0.35	-0.24	...	0.07	0.01	-0.24	-0.31	-0.14	-0.14	-0.1	-0.24	0.25	0.01	0.01	0.01	-0.07	-0.12	1		
Shore	0	0.01	0.01	...	0	0	0.01	0	...	0	0.01	-0.01	-0.01	0	0	0	0	-0.01	0.01	0.02	0	0	0	0.01	1	
New built dwelling	-0.02	0.8	-0.08	...	-0.06	-0.04	-0.01	-0.01	...	-0.01	-0.02	0.09	0.12	-0.04	-0.03	-0.05	-0.21	0.02	0.13	0.44	-0.07	0	0.21	0.21	0.01	1

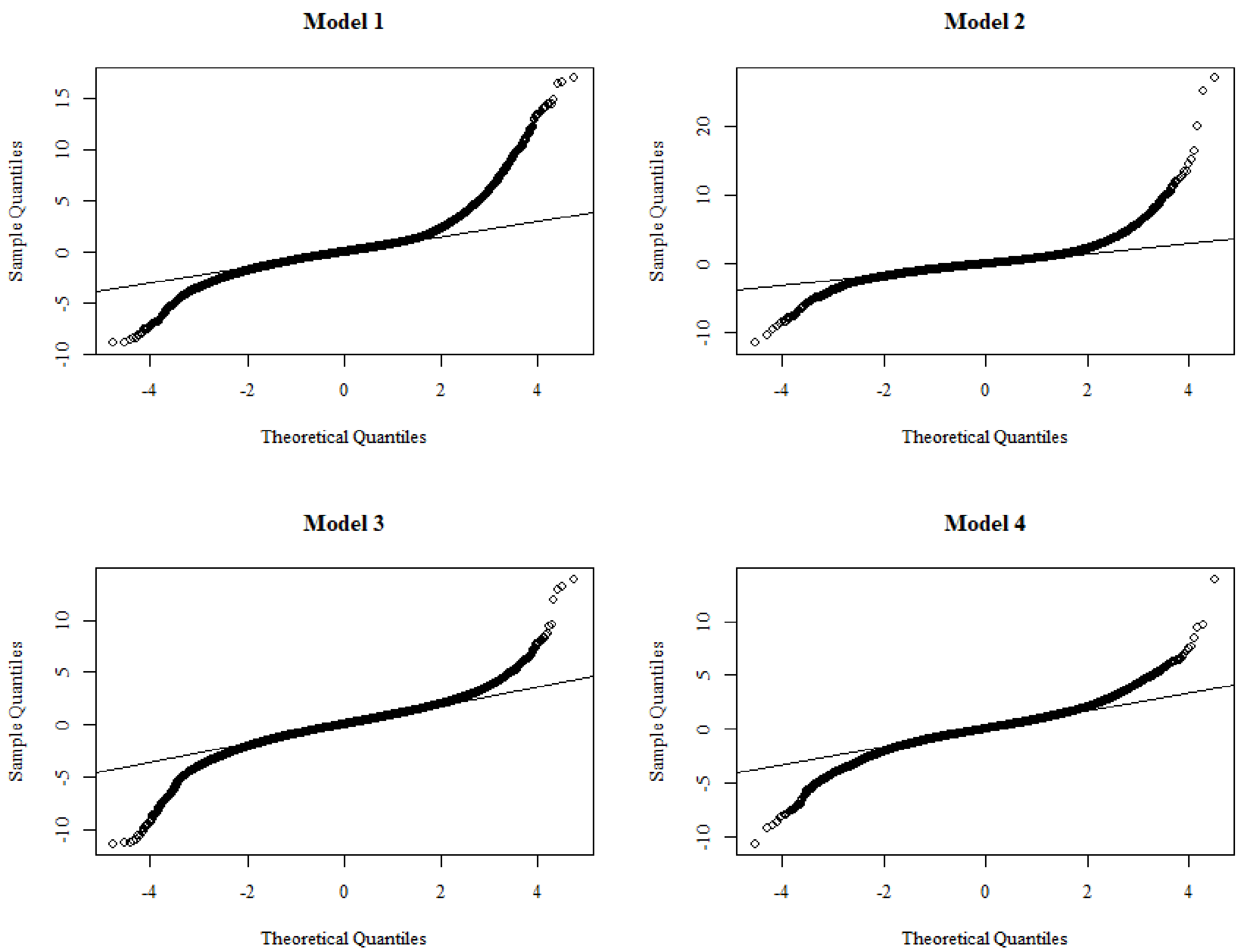
Appendix 2 - Residuals vs fitted values

Of the four models, the model (1) (unlogged dependent variable, apartment building data) has a curved shape of residuals implying non-linearity between dependent and independent variables. The model (2) (unlogged dependent variable, row house data), (3) (logged dependent variable, apartment building data) and (4) (logged dependent variable, row house data) seem to fulfill the linearity assumption. Log-models (3) and (4) also have more homoscedastic error terms than linear models (1) and (2). The residual plots suggest using the natural logarithm of the *Price per square meter* as our dependent variable. All of the residual plots have a strange looking sharp line in the lower left corner, which is caused by the *Price per square meter* cutter at 770 euros that we applied in section *4.2 – Deriving the final dataset*.



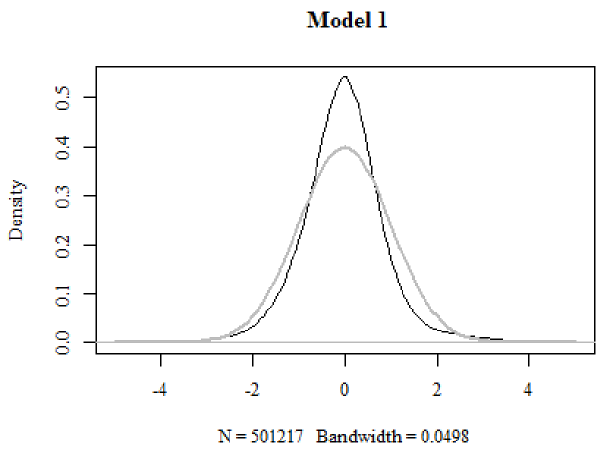
Appendix 3 – Q-Q plots of the standardized residuals

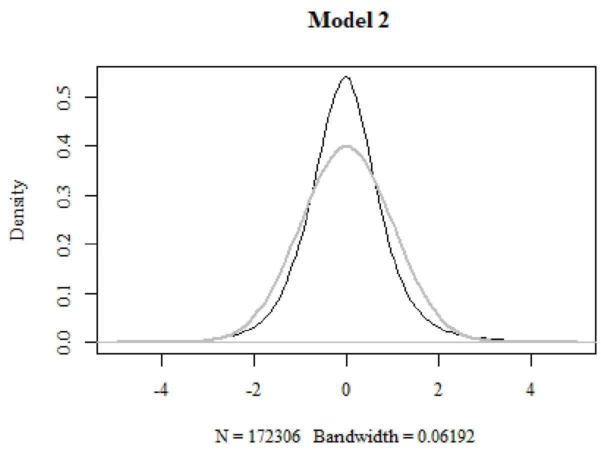
Q-Q plots tell whether the residuals are normally distributed or not. Models (1) and (2) (unlogged) perform badly forming a strong S-shaped curve. Models (3) and (4) (logged) perform better forming a relatively lesser S-shape. The curving ends of the residual plots show that we observe extreme residuals from our regressions more often than what would be expected if the residuals followed a normal distribution. The residuals have a leptokurtic distribution meaning fatter tails than a normal distribution would have. Our residuals are not exactly normally distributed, but residuals from models (3) and (4) follow the normal distribution relatively well, and thus fulfill the normality assumption of the OLS method.

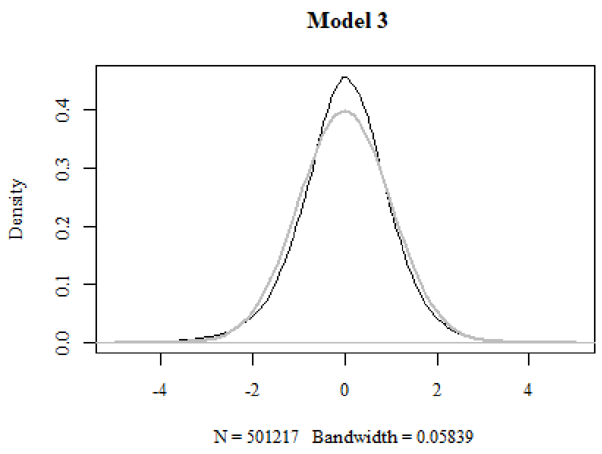


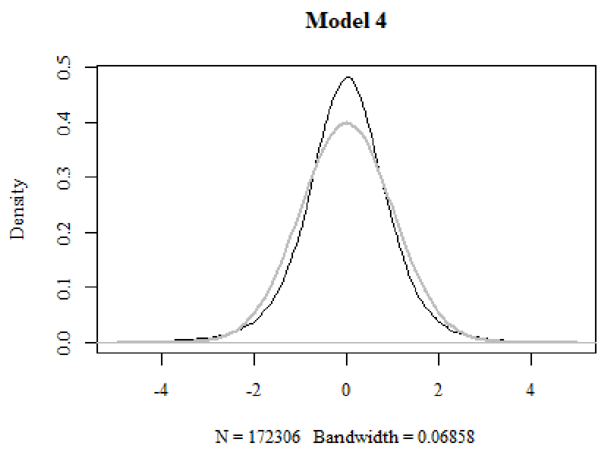
Appendix 4 – Density functions of standardized residuals

Density function plots of the standardized residuals give the same information as the Q-Q plots from a different perspective. The ideal form of normal distribution is also shown in all plots on light grey color. Our models (3) and (4) follow the normal distribution relatively well but have slightly fatter tails and bit more weight close to the mean than the normal distribution.









Appendix 5 – Model testing with tighter observation restrictions

We test how the residuals of model (3) act when applying a 10-euro cutter in *Maintenance charge per square meter* and 10 000-euro cutter in *Price per square meter*. We find that the tighter restrictions make no difference in the fit of the model. The tighter restrictions only reduce the amount of data available.